



PRESERVACIÓN DIGITAL

MIQUEL TÉRMENS



Preservación digital

Miquel Térmens

Colección *El profesional de la información*
Dirección: Javier Guallar y Tomàs Baiget
Diseño del libro y de la cubierta: Natàlia Serrano

Primera edición en lengua castellana: junio 2013
Primera edición digital: febrero 2014

Gran Via de les Corts Catalanes, 872, 3a planta
08018 Barcelona
www.editorialuoc.com

Realización editorial: Sònia Poch Masfarré
Realización digital: Sònia Poch Masfarré

ISBN: 978-84-9064-082-1

Índice de contenido

¿A QUIÉN VA DIRIGIDO ESTE LIBRO?

Este libro resultará de interés a quienes:

- Deseen conocer qué técnicas existen para preservar contenidos digitales para el futuro.
- Necesiten acceder a ficheros creados en el pasado, por ejemplo los incluidos en donaciones o legados testamentarios.
- Gestionen bibliotecas digitales o archivos electrónicos.
- Gestionen repositorios institucionales y deseen saber cómo migrar en el futuro sus datos y metadatos a otros sistemas.
- Les interese saber cómo preservar los datos en bruto de las investigaciones.
- Necesiten asegurar la pervivencia de los datos que actualmente recopilan sistemas como Europeana o la Biblioteca Digital Hispánica.

INTRODUCCIÓN

Una de las bases de la cultura y también de la economía es conservar la producción intelectual del presente para que pueda ser utilizada de nuevo en el futuro. Con la llegada de la información digital este principio se mantiene pero a costa de un mayor esfuerzo, pues los datos y los documentos digitales son más difíciles de conservar que los plasmados en soportes tradicionales como el papel.

Dependemos de la tecnología informática, pero esta cambia continuamente y por tanto es efímera. Los aparatos informáticos a medio y largo plazo se convierten en materia de exhibición en museos porque se trata de artefactos que representan la tecnología propia de una civilización, posiblemente inservibles después de su momento de uso. Pero los datos digitales no corren la misma suerte, pueden pervivir y seguir siendo usados más allá de la vida del software y del hardware empleados en su creación. Por ello no decimos que los conservamos, sino que los preservamos para que puedan volver a ser utilizados.

No es una meta fácil: no olvidemos que los datos digitales son dinámicos, no solo se pueden usar (leer, visionar, escuchar), como los datos analógicos, también se pueden manipular (interactuar, editar, borrar). Esto es lo que debemos preservar: toda su riqueza y toda su complejidad.

En este libro se muestra, en primer lugar, que la preservación de los objetos digitales no es meramente una cuestión técnica, un problema que nos haya legado la informática. Las barreras de índole organizativa, económica o legal pueden ser aún mayores que las técnicas y dificultar que las soluciones de preservación se mantengan (sean sostenibles) a medio y largo plazo.

En segundo lugar, se explica OAIS, el modelo teórico que actualmente rige el diseño de sistemas de preservación digital de todo tipo. Es importante entender correctamente su filosofía para poder dar con soluciones acordes con cada situación. Las técnicas de refresco de soportes, migración de formatos, emulación de sistemas informáticos y de análisis forense digital son

explicadas a continuación como las grandes estrategias técnicas existentes en la actualidad. A un tercer nivel, aún más técnico, se encuentra el uso de distintas herramientas y estándares para el control de los ficheros y de los metadatos asociados.

Se ha preparado un capítulo dedicado a la presentación sumaria de las soluciones aplicables a distintos escenarios sectoriales. En esta exposición no están todos los sectores económicos o de actividad con problemáticas de preservación digital (de hecho todos se ven afectados en mayor o menor medida), y tampoco se repasa la totalidad de tipos de información o documentación digital, pues ello sería imposible dadas las dimensiones de este libro. Los sectores analizados son aquellos en los que se ha manifestado una mayor preocupación por la problemática de la preservación y, por tanto, en los que con mayor facilidad puede estar implicado un profesional lector de esta obra. Dado que esta obra se publica en España, se ha considerado conveniente presentar algunos ejemplos de buenas prácticas de este país.

Finalmente adelantamos uno de los puntos clave de este libro y que será recordado en más de una página a lo largo del mismo. La problemática de la preservación digital no es la misma en todos los casos ni en todos los sectores, y tampoco en todas las empresas e instituciones de un mismo sector. En la misma línea, los ejemplos de este libro se han de entender como pistas para su adaptación a otras situaciones. Por ello, que nadie crea que comprando un determinado software o hardware habrá solucionado ya esta problemática. Por el contrario, es necesario llevar a cabo en todos los casos un análisis de las necesidades y de los medios disponibles, para encontrar luego la solución adaptada a esa realidad.

¿Qué es la preservación digital?

En el mundo actual toda la información puede ser reducida a un código binario con el que se representan números, letras, sonidos e imágenes. Todos los datos y los documentos pueden ser digitalizados y gestionados por medios informáticos gracias al software y hardware específico. Pero el avance tecnológico mantiene a la informática como una disciplina sometida a continuos cambios y mejoras.

Desgraciadamente ello provoca que el software, el hardware y los formatos de los ficheros tengan una vida limitada. La vida útil del hardware se estima en unos cinco años y a veces menos, no solo porque los nuevos equipos son mejores y más baratos, sino también debido a que los antiguos se estropean, pues están formados por componentes electrónicos con una vida útil limitada. En el caso de los sistemas de almacenamiento, como disquetes, discos ópticos y discos magnéticos, su vida también es limitada, con el agravante de que un error físico o de manipulación puede conllevar la pérdida irreversible de información.

El software está optimizado para trabajar en un determinado entorno hardware y para atender unas determinadas prestaciones requeridas por los usuarios. Cuando cambian el hardware o los requerimientos, el software se vuelve obsoleto y se acaba sustituyendo por una nueva versión del mismo o por un producto mejor lanzado por la competencia. Es cierto que un usuario puede decidir no cambiar de software porque este ya cubre perfectamente sus necesidades, pero a medio plazo también se verá obligado a deshacerse del mismo porque el hardware que lo ejecuta dejará de funcionar o de comercializarse.

Los datos se codifican y se almacenan en ficheros. Estos ficheros tienen estructuras distintas, llamadas formatos, que permiten adaptarse mejor a determinados tipos de datos y a determinados usos. Los formatos son, por tanto, una plasmación de un determinado conocimiento informático, en íntima relación con el software y el hardware con el que se van a ejecutar. Con el

paso del tiempo los formatos se vuelven obsoletos porque aparecen otros que se adaptan mejor a los cambios habidos en el software, el hardware y las prestaciones reclamadas por los usuarios.

Los cambios en el hardware, el software y los formatos son inevitables y debemos convivir con ellos. El problema aparece cuando los datos digitales tienen una vida útil que va más allá de la vida del hardware, el software y los formatos que los sustentan. Debemos de prever estos cambios e incorporarlos en la gestión de los datos; esta es la misión de la preservación digital.

En este punto es necesario conocer las diferencias existentes entre tres conceptos relacionados:

- Conservación. Pone el énfasis en establecer unas condiciones ambientales, de almacenamiento y de uso que permitan conservar un documento en las mejores condiciones posibles de estado físico y de uso. Así, por ejemplo, se preocupa de que los niveles de humedad y temperatura sean los correctos, que la manipulación de los documentos durante su utilización no los deteriore, que el almacenamiento facilite el mantenimiento del estado mecánico de los soportes documentales, etc. Las técnicas de conservación son claves en el cuidado de los documentos analógicos, pero también son de aplicación con los digitales. Por ejemplo, un disco magnético puede ser destruido por la acción del fuego igual que un pergamino, por lo que los sistemas de prevención de incendios son de aplicación obligatoria en los dos casos.
- Seguridad informática. Establece políticas para el análisis, la detección y la posible solución de los riesgos de tipo informático que pueden sufrir los datos en un determinado sistema informático. Se preocupa de establecer procedimientos de copias de seguridad, de controlar el acceso autorizado a los datos, de analizar la fiabilidad del hardware y el software utilizado, de detectar y neutralizar los ataques de virus o de piratas informáticos, de prever contingencias debidas a cortes del suministro eléctrico, etc. Su actuación se centra en el presente y en el corto plazo, y normalmente está ligada a los ámbitos de gestión, al día a día. Así, que un centro de datos esté certificado según la norma ISO 27001:2005 indica que dispone de un alto nivel de seguridad informática, pero no necesariamente que sus datos aún sean legibles dentro de 30 años. Aunque la seguridad informática se caracteriza por sus actuaciones tecnológicas, no olvida que la seguridad integral también depende de aplicar de forma

adecuada medidas organizativas (por ejemplo, quién es el responsable de los datos), ambientales (una inundación puede destruir un centro de datos) y formativas (el personal ha de mantener el secreto de los códigos de acceso), entre otras.

- Preservación digital. Asegura el acceso y el uso futuro de los documentos digitales creados en el presente o el pasado. A partir de las políticas de conservación y de seguridad informática, añade otras (migración, emulación...) que permitan su mantenimiento y uso a largo plazo.

La conservación pone el énfasis en los aspectos preventivos y pasivos de la salvaguardia documental y la seguridad informática se centra, aunque no únicamente, en los aspectos tecnológicos. La preservación digital se nutre de estas visiones previas y las amplía para que puedan surtir efecto a largo plazo.

Previamente hemos observado que los datos digitales viven en un entorno inestable debido a los cambios en el hardware, el software y los formatos, pero aquí no acaban sus problemas. Debemos añadir que los datos digitales son muy frágiles: fácilmente se pueden cambiar o borrar, sea de forma consciente (yo edito un documento), por error (yo edito una versión equivocada del documento), por negligencia (no he hecho la copia de seguridad) o por una acción externa (un virus me ha alterado el documento).

Esta fragilidad es muy evidente durante el tiempo de creación y de uso habitual de los datos, pero lo sigue siendo a lo largo de su vida. En cualquier momento, ahora o dentro de 30 años, un virus podría corromper la base de datos de la Tesorería General de la Seguridad Social y dejar al lector sin su derecho a obtener una pensión de jubilación. La fragilidad conlleva otro problema: el cuestionamiento de la integridad y la autenticidad del fichero.

Además de depender de la fragilidad de los datos, de sufrir problemas para asegurar su autenticidad, nos encontramos también con el carácter irreversible de su posible pérdida. Así, un insecto o un roedor pueden provocar un agujero en una fotografía en papel y ello afectar a un detalle de la imagen, pero el resto de la fotografía continuará siendo igual de visible y la propia autenticidad del documento no se verá alterada en absoluto por este contratiempo. En cambio, un virus o un error de grabación de un fichero quizás solamente afecten a 2 bytes de una fotografía digital, pero al corromper el algoritmo del formato pueden dejar el fichero totalmente indescifrable, sin que los programas previstos lo puedan mostrar. En el entorno analógico una restauración puede

paliar o incluso solucionar los problemas de conservación de un documento en papel, pero la restauración no existe en documentos informáticos, como máximo se pueden aplicar técnicas de recuperación parcial de datos, no de restitución de su estado original.

La preservación digital se basa en la aplicación de técnicas activas de conservación informática porque sabemos que ante la aparición de un problema quizás ya no podremos dar con una solución; ante un borrado de datos no hay restauración posible. Por esta misma razón, no es posible aplicar políticas de negligencia benigna en el entorno digital. Ante documentos analógicos podemos optar por no hacer nada, dejarlos encerrados en un depósito y esperar que al cabo de años la falta de interacción humana no habrá podido provocar ningún tipo de pérdida, de forma que los documentos posiblemente aún se encontrarán en un buen estado de conservación. Si con los datos digitales también somos negligentes, por ejemplo abandonando un disquete dentro de un archivador durante 40 años, solo conseguiremos perder su contenido, pues al cabo de este tiempo este soporte ya no será legible o ya no existirán los medios técnicos apropiados para leerlo.

CÓMO ALCANZAR LA PRESERVACIÓN DE LOS OBJETOS DIGITALES

Aunque la preservación digital es la respuesta a un problema de origen tecnológico, no se puede alcanzar aplicando solamente medidas técnicas, que de todas formas se tratarán en este libro más adelante. También es imprescindible que se tomen medidas a nivel organizativo, financiero y legal. Adelantamos ya que este tipo de soluciones pueden resultar de más difícil aplicación que las simplemente tecnológicas y normalmente están más allá de las competencias del personal especializado en preservación.

Aspectos organizativos

El primer paso en una iniciativa de preservación es determinar que una información digital debe ser preservada y el segundo concretar quién es el responsable de esta misión.

Recordemos que a lo largo de la vida de una información o de un documento, estos pasan por las manos de distintos actores que tienen intereses diferentes en los mismos. Estos actores pueden ser, como mínimo: quien creó el documento, quien lo usó, quien lo gestionó durante su vida útil, quien lo preserva a largo plazo y quien lo va a usar (quizás) en el futuro. Dichos actores pueden coincidir, pero no siempre los que han tenido intereses en un documento en el pasado (incluso por haberlo creado y explotado) desean asumir la responsabilidad y los costes de su preservación a largo plazo. Un ejemplo lo tenemos en muchas publicaciones, como la prensa digital, cuya explotación está básicamente en manos privadas, con un modelo financiero basado en las ventas y la publicidad del presente; unas empresas editoras que no siempre están comprometidas con el sostenimiento a largo plazo de unos diarios que, en el futuro, siempre serán de gran interés para conocer el pasado de un país.

Incluso pueden existir problemas en el caso de que un actor responsable de la preservación esté plenamente convencido de la necesidad de preservar

determinada documentación. Aquí las dificultades pueden aparecer en el seno de la organización, donde no siempre es fácil delimitar a quién corresponde la responsabilidad: ¿a la unidad que gestionó inicialmente los documentos, al servicio de informática, al archivo, a la alta dirección, o a una combinación de los anteriores?

Si la preservación es una actividad que genera valor dentro de una organización, es muy probable que aparezcan disputas entre departamentos para aparecer como los responsables de la misma. Si por el contrario la preservación no genera valor y se debe realizar por obligación legal, patrimonial o de otro tipo, internamente será vista como una carga para la organización. En este caso se resaltarán las dificultades técnicas y los costes; los departamentos más bien tenderán a hacerse los desentendidos y a no dar soporte.

Ejemplos del primer escenario son el archivo de análisis clínicos de un laboratorio farmacéutico y el archivo de producción propia de una emisora de televisión. Ejemplos del segundo escenario son los archivos intermedios de algunas administraciones públicas.

Costes y financiación

En la actualidad el factor económico es el que tiene más incidencia como freno de las actividades de preservación digital. A corto plazo los problemas en este ámbito se centran en obtener la adecuada financiación, pero a medio y largo plazo se les suma el aumento de los costes que provoca el aumento de las colecciones a tratar.

Las dificultades para delimitar las responsabilidades sobre la preservación de unos documentos conllevan normalmente la aparición de obstáculos para conseguir una financiación adecuada para estas actividades. Así pues, los problemas organizativos y los financieros están directamente relacionados. No olvidemos tampoco que se trata de actividades caras, como otras inversiones relacionadas con la informática, y que, para colmo, no generan un rendimiento tangible inmediato: preservar ficheros no ayuda a ganar elecciones y a menudo tampoco es capaz de generar ganancias económicas.

Como toda actividad nueva, es normal que la preservación se inicie bajo la forma de proyectos de innovación soportados, por tanto, con presupuestos extraordinarios, incluso bajo la forma de subvenciones públicas. Pero más

adelante las actividades de preservación digital han de pasar a ser consideradas como gastos corrientes, como la electricidad, la limpieza de los edificios o la realización de copias de seguridad de las bases de datos, solo así se podrá asegurar su adecuada financiación, aun en épocas de dificultades económicas.

La estructura de los costes es bien conocida por los expertos pero no es la que cree conocer el público no especializado porque no todas las actividades involucradas en un sistema de preservación digital tienen el mismo nivel de costes y estos a veces no son los más evidentes. Aunque los gastos de cada proyecto pueden ser diferentes, una aproximación global es la siguiente:

- Ingestión de datos en el sistema: 50 % de los costes totales.
- Acciones propiamente de preservación (como almacenamiento de datos y migración de formatos): 33 %.
- Acceso a los datos: 17 %.

Por tanto a nivel funcional, y esto es lo importante, la mayoría de los gastos se producen en la ingestión de los datos porque se trata de las tareas que consumen más personal y en las que con más facilidad se pueden dar problemas en la automatización debido a incidencias con los ficheros entrantes.

Los costes de almacenamiento están formados básicamente por los costes de compra de disco duro, el medio de almacenamiento que actualmente se usa de forma predominante. Como acabamos de indicar y en contra de la opinión comúnmente aceptada, no son la mayor fuente de gastos, aunque sí se convierten en un gran gasto permanente. Hasta ahora estos discos han ido reduciendo su coste unitario año tras año siguiendo lo que se conoce como ley de Kryder: la densidad de almacenamiento en los sistemas magnéticos se dobla cada aproximadamente 13 meses manteniendo el mismo coste total, lo que significa que los costes de almacenamiento se reducen a la mitad cada 13 meses.

De seguir así, la necesidad futura de almacenar más datos no supondría un gran reto porque su coste unitario iría bajando de forma paralela a los requerimientos de mayor almacenamiento. Algunos expertos creen que esta tendencia ya está cambiando y no se va a mantener en el futuro porque el ritmo de innovación técnica en discos magnéticos se está frenando, lo que comporta que los precios por gigabyte almacenado no se están reduciendo al mismo ritmo que antes. La resolución de esta incógnita en uno u otro sentido a medio

plazo puede obligar a cambiar los presupuestos necesarios para preservación digital y, con ello, quizás también las estrategias seguidas. Si, por ejemplo, crece la necesidad de recortar los gastos en almacenamiento ello puede obligar a la adopción de formatos con mayor capacidad de compresión y, por tanto, que generen ficheros más pequeños, como es el caso de la adopción del formato JPEG2000 en detrimento del formato TIFF, como ya han hecho algunas grandes bibliotecas. Seguro que en el futuro van a crecer las tensiones para elegir entre formatos robustos a nivel técnico u otros formatos más económicos debido a su menor uso de almacenamiento.

Además del incremento de las necesidades de almacenamiento, otros dos retos amenazan los repositorios de preservación: los costes de personal y el aumento exponencial del número de objetos a tratar en la fase de ingestión. Para hacerles frente una primera línea de actuación consiste en automatizar al máximo los procedimientos de trabajo, en especial en la fase de ingestión, para poder aumentar el número de objetos gestionados sin aumentar el personal destinado a los mismos. Una segunda línea de actuación es establecer claras políticas de selección de los objetos a conservar con el fin de limitarlos al número mínimo imprescindible. Aunque esta parece ser una línea de actuación lógica, pues forma parte de los principios de cualquier gestión documental clásica, en la práctica en muchos casos no es posible aplicarla, pues los costes de selección pueden ser mayores que los costes de preservar material innecesario. Esta elección realizada bajo principios racionales de análisis de costes tendrá una consecuencia inesperada: llegarán al futuro datos y documentos que no van a interesar a nadie (¿o sí?): duplicados, descartados, truncados, borradores, pruebas...

Aspectos legales

Actualmente aún no se encuentran resueltos todos los aspectos jurídicos de la preservación digital. Las legislaciones nacionales e internacionales de protección de los derechos de autor impiden cambiar el soporte y el formato de los documentos sin la autorización expresa de quien disponga de los derechos de propiedad intelectual. Cambiar los contenidos de un soporte a otro (por ejemplo de un CD-ROM a un disco duro) o migrar los ficheros de un formato obsoleto a otro actual son actuaciones habituales en un sistema de preservación digital. En algunos casos puede ser necesario superar el bloqueo

de una protección DRM o aplicar ingeniería inversa a un software con el fin de analizar su funcionamiento y poder conservarlo; todas estas acciones también están prohibidas por las regulaciones de propiedad intelectual y de patentes, que las asimilan a acciones de pirateo. Es urgente, por tanto, que estas restricciones puedan ser salvadas cuando se justifique que se realizan con finalidades de preservación digital, pero para ello es necesario que se convenza a los legisladores para que establezcan excepciones legales favorables a la preservación digital en determinados contextos. Tanto en Estados Unidos como ante la Unión Europea se han presentado estudios y propuestas concretas para eliminar, aunque sea parcialmente, estas barreras. Pero en último término estas propuestas aún no han conseguido triunfar debido a que siempre han prevalecido los miedos esgrimidos por las entidades defensoras de los derechos de autor, que han visto las posibles excepciones por causas de preservación digital como una brecha peligrosa en sus derechos.

Otro tema legal, pero de naturaleza distinta, que debe tenerse bien en cuenta es la protección de los contenidos preservados ante usos no autorizados. Según el tipo de ficheros conservados, nos podemos encontrar ante datos de carácter personal, datos financieros de personas o de empresas, datos que pueden afectar a la imagen personal, datos médicos, datos susceptibles de ser patentados, etc. Estas afectaciones pueden ser permanentes o, más comúnmente, variar con el paso de los años. Recordemos que el paso del tiempo también puede producir cambios en la propia naturaleza del servicio de preservación, que en el presente sirve a una institución de un país determinado, con una legislación determinada, pero en el futuro puede adoptar una meta más amplia, con más legislaciones aplicables. Un ejemplo lo tenemos en la base de datos de libros HathiTrust, que para la identificación de los derechos de autor de un libro prevé 19 Estados distintos y 17 fuentes de los mismos, lo que generará posibilidades de consulta distinta en Estados Unidos o en otros países.

Una preservación sostenible y fiable

Si pueden parecer difíciles de resolver los aspectos organizativos, económicos y legales que acabamos de repasar, aún lo serán más si tenemos en cuenta que nos movemos en escenarios cambiantes. Los sistemas de preservación digital, en una primera toma de contacto, recuerdan a archivos

históricos o a procedimientos de copias de seguridad y, por tanto, a sistemas inamovibles, ajenos a los cambios que se producen en el resto de la sociedad. Esta es una impresión totalmente equivocada porque la mayoría de sistemas de preservación digital están pensados para crecer sin fin, como también lo hace la producción o las necesidades de los organismos de los que dependen. Además, con el paso del tiempo cambian las necesidades de los actores implicados (productores, gestores, usuarios...), la naturaleza técnica de los ficheros debido a la evolución de la tecnología informática, la disponibilidad financiera para sostener las actividades de preservación digital y la valoración que hacemos de las mismas.

Los documentos digitales son frágiles y las técnicas para preservarlos también. Esta fragilidad de la preservación digital no viene dada tanto por estar sustentada en metodologías y técnicas complejas y aún no del todo probadas, como por el hecho de que todas las actuaciones pueden verse comprometidas por una sola mala acción: una negligencia, una mala formación del personal, un fallo técnico o... un recorte en los fondos económicos que la sustentan. Todos estos problemas técnicos, organizativos, financieros y legales también se han de prever y se deben planificar las medidas correctoras que faciliten su gestión. La preservación debe ser sostenible a nivel técnico, organizativo, económico y legal, y esta sostenibilidad se ha de poder demostrar a todos los que estén implicados: la administración, las entidades que aportan la financiación, los propietarios de los contenidos, las comunidades de usuarios. Gracias a esta demostración de fiabilidad se podrá confiar en el sistema de preservación. En el caso de instituciones que preservan datos de terceros es aún más necesario obtener confianza ante estos terceros mediante el desarrollo de una buena actuación en preservación y poder demostrarla externamente.

Las auditorías son un mecanismo reconocido para testar determinado servicio mediante un examen profesional por parte de expertos independientes, lo que dota de credibilidad a su dictamen. En sistemas informáticos las auditorías son una práctica habitual y por ello también deberían ser de aplicación obligatoria en los sistemas de preservación digital. Los actuales sistemas de gestión de la seguridad parten de las recomendaciones de la norma británica BS 7799, después reconvertida en la familia de normas internacionales ISO 27000, conocidas en España con la nomenclatura UNE-ISO/IEC 27001:2007

y UNE-ISO/IEC 27002:2009. Estas normas utilizan la misma metodología de mejora continua —el modelo PDCA (*Plan-Do-Check-Act*)— usada en las normas de gestión de la calidad (ISO 9000) y de gestión medioambiental (ISO 14000). La norma ISO 27001 permite que un sistema de información pueda ser auditado por un equipo externo y como resultado pueda obtener una certificación acreditativa.

En el caso de las administraciones públicas españolas, el Esquema Nacional de Seguridad (ENS) ha establecido mediante el Real Decreto 3/2010 la obligatoriedad de disponer de un sistema de gestión de la seguridad basado en el análisis de riesgos y también ha normalizado un sistema propio de auditorías.

Los sistemas especializados en preservación digital presentan particularidades de planificación y gestión que no son revisadas con el detalle necesario en las normas que se han mencionado previamente. Por esta razón, en diversos países distintas instituciones y grupos de interés han formulado sistemas de auditoría especializados en el ámbito de la preservación digital, entre los que destacan:

- Drambora (Reino Unido): http://www.repositoryaudit.eu/
- Nestor 2 (Alemania): http://files.d-nb.de/nestor/materialien/nestor_mat_08_eng.pdf
- TRAC (Estados Unidos): http://public.ccsds.org/publications/archive/652x0m1.pdf. Este último ha sido finalmente reconocido como norma ISO/IEC 16363:2012 *Audit and certification of trustworthy digital repositories* y se está organizando un cuerpo de auditores especializados en su aplicación.

La extensión de los sistemas de auditoría sin duda obligará a aumentar el rigor con el que se gestionan los centros de preservación digital pero por sí mismos no resolverán el mayor problema que amenaza la persistencia de estos centros, que es la sostenibilidad económica o, dicho en otras palabras, asegurar una financiación correcta a largo plazo que permita ejecutar correctamente sus actividades al mismo tiempo que aumenta el número de fondos que deben conservar.

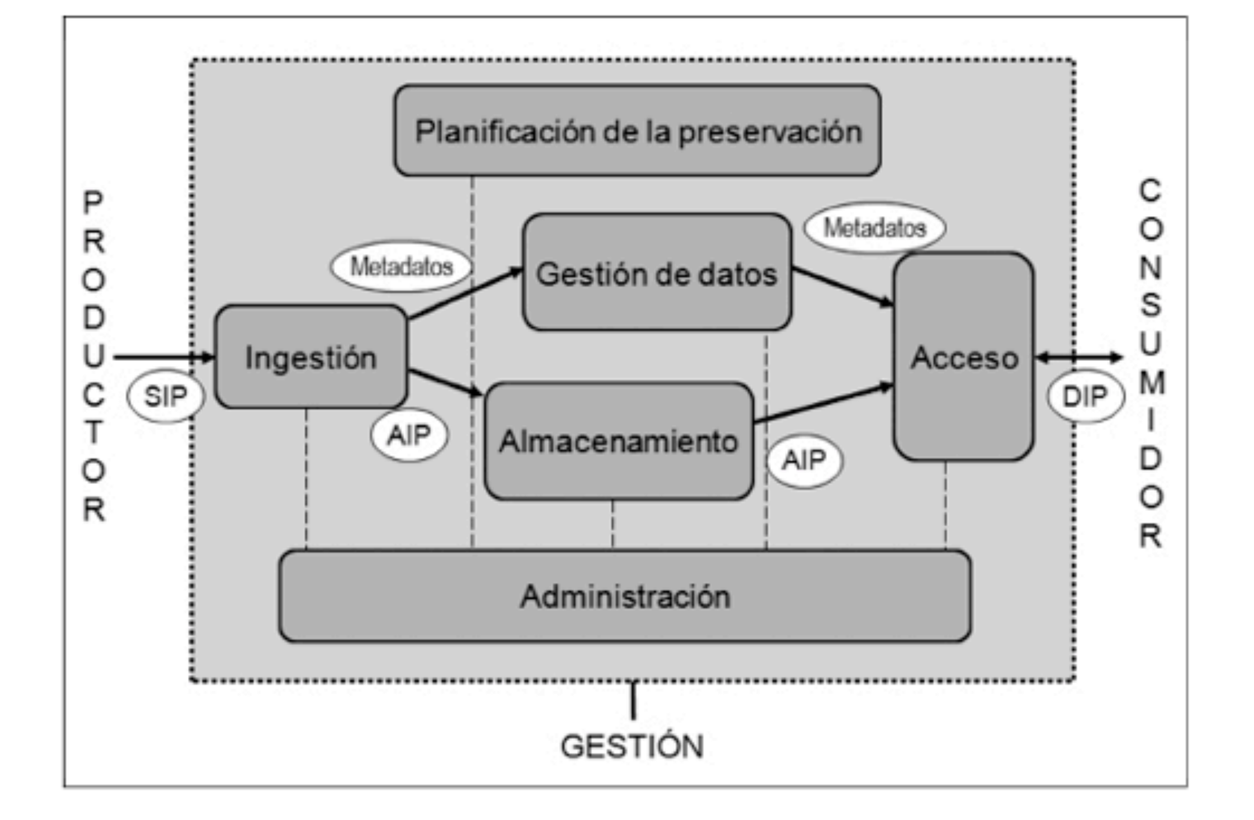
El modelo OAIS

En la década de los noventa del siglo xx la gran pregunta era cómo debía ser un sistema de preservación digital a nivel técnico, no tanto en su concreta configuración de hardware y software, sino a nivel de las características que debía tener y las funciones que debía cumplir. Era necesario inventar una nueva categoría de aplicación informática de la que se conocían los problemas que tenía que resolver, pero no cómo. La solución llegó del trabajo de la NASA.

La NASA es un organismo cuyo funcionamiento depende en gran medida de la reutilización de datos conservados del pasado, por ejemplo los valores de telemetría de los distintos vuelos al espacio. Desgraciadamente la NASA ha confirmado la pérdida de datos del pasado, como las imágenes originales en vídeo del primer alunizaje del hombre a la Luna, el del Apolo 11 en 1969, del que solo se conservan vídeos derivados de menor calidad, o los datos de las primeras muestras de suelo tomadas el año 1975 en la superficie de Marte por las sondas Viking.

Ante estos errores en la custodia de datos únicos, la NASA tomó conciencia de que la preservación a largo plazo de datos digitales se tenía que abordar desde planteamientos distintos a los empleados hasta ese momento. El resultado fue el desarrollo de un modelo teórico que integrase y explicase las funciones que debería cumplir cualquier sistema integral de preservación. El modelo fue discutido y aprobado dentro del Council of the Consultative Committee for Space Data Systems (CCSDS), el organismo encargado de desarrollar los estándares de datos de las principales agencias del espacio a nivel mundial, tomando el nombre de *Reference Model for an Open Archival Information System (OAIS)*, publicado en enero de 2003. OAIS se convirtió en norma ISO el año siguiente con el código 14721:2003. Posteriormente, en junio de 2012, el CCSDS publicó una revisión que no incorporó cambios sustanciales (http://public.ccsds.org/publications/archive/650x0m2.pdf).

Ilustración 1. Esquema funcional del modelo OAIS



OAIS es un modelo teórico que indica qué funciones han de soportar los sistemas de preservación digital, sin importar qué tipo de datos custodian ni a qué tipo de actividad u organización se refieren. No es por tanto un software, un hardware, un formato o unas normas de codificación. OAIS describe seis grandes bloques de procesos dentro de un archivo de preservación digital (véase la Ilustración 1):

1. Ingesta o ingestión. Los ficheros llegan procedentes de los productores (las oficinas, los sistemas informáticos de gestión, la captura de información de la red, etc.) y se les aplican una serie de controles antes de proceder a su ingreso en el sistema de preservación. Algunos controles son:

a) Control de procedencia e integridad de la remesa: no faltan ni sobran ficheros y estos no se han corrompido o alterado desde su punto de envío.

b) Control antivirus: los ficheros no contienen virus.

c) Control de formatos: identificar de forma clara el formato y la versión de cada fichero y si este está bien formado.

Después de estos controles el sistema extrae metadatos de tipo técnico de los ficheros (resolución, número de colores, número de palabras, codificación de caracteres, etc.) y crea una firma digital o *checksum* de los mismos, con el fin de que en el futuro se pueda verificar su integridad. Finalmente los ficheros, junto con su valor *checksum*, son enviados al proceso de *Almacenaje*, mientras

que los metadatos (los que ya incorporaba el fichero, más los técnicos creados en la ingesta, más datos de antivirus, etc.) son enviados a *Gestión de datos*.

2. Almacenaje. Es el proceso encargado de almacenar físicamente los ficheros de datos; por tanto está formado por *racks* de discos magnéticos u otros sistemas de almacenaje de alta fiabilidad, controlados con protocolos de copia de seguridad y de redundancia de datos (ver «Cambio de soporte», más adelante).

3. Gestión de datos. En este proceso se mantienen los metadatos de los ficheros: los originales, los creados durante la ingesta y los que se van generando a lo largo de la vida de los ficheros. El objetivo es disponer de forma centralizada y normalizada de todas las informaciones que puedan facilitar la conservación y el uso de cada fichero; en este sentido es importante registrar todas las incidencias que sufra a lo largo del tiempo, como pueden ser migraciones de formatos, alteraciones en la integridad, resellado de tiempo y firma electrónica. El formato de metadatos PREMIS ha sido creado para registrar este tipo de informaciones, aunque se pueden usar otros mecanismos de control.

4. Acceso. Se han de habilitar procedimientos que permitan el acceso de los usuarios a los contenidos preservados. *Acceso* ha de disponer de algún tipo de interfaz que permita la interrogación de los metadatos custodiados en *Gestión de datos* y, a partir de sus resultados, dar acceso los contenidos que se encuentran en *Almacenaje*. Aquí es preciso recordar dos puntos importantes:

a) Que determinados datos y documentos hayan sido preservados hasta el futuro no implica que vayan a ser de acceso libre. Por esta razón en *Acceso* se deberán integrar las políticas de identificación de usuarios y de derechos de acceso a los contenidos que sean pertinentes.

b) Los formatos en los que el usuario del futuro va a consultar un contenido no tienen por qué ser los mismos formatos en que fueron creados en el pasado; es más: a medio y largo plazo los contenidos que habrán migrado de formato serán la mayoría. Así, serán distintos el formato original, el formato de preservación que está almacenado y el formato de acceso en un momento determinado (un ejemplo puede ser un texto que se creó con WordPerfect 5.1, que se migró a y se está preservando como Microsoft Word 97 y que en el futuro va a ser consultado como Libre Office versión 2050). A nivel técnico, *Acceso* deberá disponer de los mecanismos para migrar los formatos bajo demanda

(según las posibilidades de visualización de los usuarios) y proporcionar, si es necesario, los visores o el software cliente para abrir los ficheros. *Acceso* también deberá informar a los usuarios de las salvaguardas de uso (derechos de propiedad intelectual, protección de datos personales, informaciones confidenciales) aplicables a un objeto digital determinado.

5. Preservación. Es la parte inteligente del sistema, aquí es donde se deciden las políticas a aplicar. Los responsables de un sistema de preservación han de mantener una vigilancia tecnológica que los alerte del fin del tiempo de vida de un formato y de la necesidad de migrar sus ficheros a otro formato (ver más adelante: *Migración de formatos*), de problemas en la operatividad de un formato o software, de la disponibilidad de nuevas herramientas de visualización o de emulación, etc. Desde *Preservación* también se habrá de prever la propia actualización o migración del sistema de preservación, pues este también está compuesto por software y hardware que se vuelve obsoleto y que se ha de reemplazar.

6. Servicios comunes. Este proceso, como su nombre indica, es de soporte técnico a los anteriores.

OAIS también ha determinado cómo se mueven los datos entre cada uno de los bloques de procesos; ello es importante porque a menudo estos bloques pueden estar constituidos por sistemas informáticos diferenciados, incluso situados en ubicaciones físicas distintas.

- SIP o Paquete de Información Enviada (Submission Information Package). Este paquete incluye los ficheros de datos que se envían a un sistema de preservación, acompañados de aquellos metadatos que puedan ser útiles para comprobar la integridad y la autenticidad de estos ficheros; es habitual que estos metadatos de acompañamiento estén constituidos por un listado de los ficheros y de los directorios en que están organizados, así como del valor de *checksum* de cada fichero.
- AIP o Paquete de Información de Archivo (Archival Information Package). Este tipo de paquete tiene una composición y una función parecidas a las del anterior, pero aplicadas a las comunicaciones entre los bloques de Ingesta y de Almacenaje, con el fin de asegurar que los ficheros validados en la ingesta realmente son los mismos que se almacenan a largo plazo.

- DIP o Paquete de Información de Diseminación (Dissemination Information Package). Incluye los ficheros de datos que se entregan a un usuario determinado como resultado de una petición de consulta; van acompañados de metadatos para informar de la autenticidad del envío y quizás otros sobre la historia de los datos: cuándo fueron ingresados, bajo qué formato original y cuándo fueron migrados o emulados. Es posible que también se entregue alguna advertencia sobre los usos permitidos sobre estos datos (debido a restricciones de propiedad intelectual, de licencias o de protección de datos personales).

Los SIP se pueden generar de forma automática con distintos programas, como Bag-it que se usa en Estados Unidos (véase la Ilustración 2), y se pueden enviar al destinatario bajo la forma de un fichero comprimido zip, tar o equivalente. En el caso de envíos masivos, las entregas se suelen hacer con discos duros externos.

Ilustración 2. Ejemplo de paquete SIP creado con el software Bag-it

envio_001.zip
bag-info.txt
bagit.txt
manifest.md5.txt
tagmanifest-md5.txt
data
fichero1.abc
fichero2.abc
etc.

El modelo OAIS ha tenido una gran aceptación a nivel mundial, como lo demuestra el hecho de que todos los sistemas de preservación actualmente en funcionamiento o en proyecto dicen que cumplen con el modelo OAIS con mayor o menor fidelidad. Es importante recordar que las especificaciones de OAIS se han de adaptar a las peculiaridades de cada caso real. Por ejemplo, se ha advertido que la adaptación directa de OAIS a instituciones y empresas de

pequeño y mediano tamaño es farragosa, siendo necesario simplificar los procesos y las interacciones que propone OAIS. En cualquier caso OAIS es una buena guía para analizar las necesidades funcionales de cada caso, aunque quizás no tanto para dar con las soluciones concretas a implementar.

LAS TÉCNICAS

Los sistemas de preservación digital trabajan con grandes volúmenes de datos que hacen inviable un tratamiento o un seguimiento individualizado de los mismos. Las prospecciones realizadas en distintos países muestran que los mayores gastos en los sistemas de preservación son los de personal y que estos no pueden escalar al mismo ritmo que crecen los contenidos a conservar, básicamente porque resulta difícil aumentar la financiación recibida para realizar estas actividades. Ello lleva a la necesidad de automatizar al máximo todas las operaciones y a reducir la variabilidad técnica de los documentos a fin de poder crear procedimientos de gestión claros y genéricos. De aquí la necesidad de reforzar las actividades de preservación digital con el uso de estándares que aumenten la eficiencia del sistema, empezando por los que ya se aplican en el momento de la creación de los contenidos. Además, en la última década se han creado diversos estándares que vienen a normalizar aspectos esenciales de los sistemas de preservación, sea a nivel de ficheros o a nivel de metadatos. Veamos a continuación estas técnicas y estándares propios de la preservación digital.

Refresco de soportes

La fase más primaria de la preservación digital empieza asegurando que los bits no se alteran con el paso del tiempo. Pero en contra de este objetivo los soportes de almacenamiento (disco magnético, disco óptico, cinta...) presentan tres tipos de problemas:

- Los soportes tienen una vida útil limitada; incluso si han sido almacenados en óptimas condiciones de almacenamiento. Como resultado, la información grabada sufre alteraciones a nivel de bit o, en el caso extremo, se llega a la imposibilidad de poder leer el soporte.
- Los soportes se vuelven obsoletos debido a la evolución tecnológica o por razones comerciales, resultando cada vez más difícil disponer de disposi-

tivos para su lectura. Algunos soportes ya obsoletos son las cintas magnéticas para audio, los disquetes de 5 ¼", los disquetes de 3 ½" y los discos Zip.
- Los soportes pueden sufrir alteraciones físicas por causas internas o externas. Algunos ejemplos son: la acción de los campos magnéticos sobre los discos magnéticos, la humedad y los cambios de temperatura, la degradación de los componentes (plástico, laca...) de los discos ópticos, los cambios de tensión eléctrica, etc.

Para sortear el primer problema es necesario utilizar soportes de calidad (no se puede esperar la misma calidad y duración de un CD-ROM de una marca reconocida que de otro sin marca comprado en una tienda de bajo coste), así como migrar los datos de un soporte a otro antes de llegar a los límites de la vida útil recomendada por el fabricante. Estas medidas se han de reforzar realizando réplicas o copias de seguridad de los datos; no olvidemos que las cifras de esperanza de vida útil de un soporte determinado no son más que medias obtenidas en ensayos y, por tanto, se pueden dar casos de duración más corta y de duración más larga. Por otro lado, los ensayos técnicos realizados de forma independiente tienden a rebajar las cifras de esperanza de vida útil que dan los fabricantes.

Para evitar el segundo problema, la obsolescencia de los soportes, no queda otro remedio que mover los ficheros a un soporte más actual tan pronto como se sospeche que el soporte original dejará de tener aceptación. Los periodos para ejecutar un cambio de soporte son largos pero no es aconsejable apurarlos. Por ejemplo, actualmente ya no se venden disqueteras de 5 ¼" (estos discos ya no se fabrican y solo se pueden leer desde equipos viejos o usando controladoras especiales); las de 3 ½" sí, pero por poco tiempo (ya no se instalan en equipos nuevos, pero aún se pueden comprar por separado). Por tanto aún se está a tiempo de migrar los datos de los disquetes de 3 ½" sin depender de un servicio técnico externo, pero posiblemente esto ya no será posible dentro de 5 años.

El tercer problema, las alteraciones en los soportes, exige de nuevo utilizar soportes y equipos de calidad y tomar todas las medidas necesarias para minimizar los riesgos: instalación de sistemas de alimentación eléctrica ininterrumpida (SAI), control de la tensión eléctrica, aislamiento magnético, control de temperatura y humedad, etc. En el caso de almacenamientos masivos

en disco duro magnético, se recomienda usar discos de distintas marcas y pertenecientes a distintas series de producción; así es más difícil que acabemos usando discos de fabricación defectuosa.

Debe recordarse que puede ser muy elevado el total de tiempo necesario para cambiar de soporte grandes volúmenes de datos, lo que obliga a planificar de forma meticulosa estos cambios. Este es justamente uno de los principales cuellos de botella que se dan en la gestión de grandes bancos de datos, como los del CERN, donde el copiado de la totalidad de sus datos de un sistema de discos a otro más moderno puede requerir un año y medio de trabajo intensivo de varios operarios.

En resumen, si deseamos asegurar la pervivencia de nuestros datos, una institución ha de:

- Utilizar soportes y equipos de calidad.
- Realizar un adecuado uso y mantenimiento de los mismos.
- Tomar medidas para minimizar los riesgos ambientales y técnicos.
- Establecer una estrategia rigurosa de copias de seguridad.
- Cuando sea necesario, migrar a soportes de almacenamiento más modernos o con más aceptación, sin apurar los plazos disponibles para hacerlo.

Los atentados del 11 de septiembre de 2001 en Estados Unidos acabaron por alertar y convencer a las administraciones y empresas sobre los peligros de no disponer de sistemas informáticos redundantes en más de una localización. Algunas de las empresas que estaban alojadas en las Torres Gemelas de Nueva York pudieron reanudar sus operaciones al cabo de pocas horas gracias a los sistemas redundantes (hardware + software + datos + procedimientos) que habían establecido en otras localizaciones geográficas. A partir de ese suceso esta práctica ha quedado establecida como imprescindible para una correcta conservación de datos valiosos. Hay que aclarar que la replicación en más de una localización no solo se planea para prevenir ataques terroristas; también se hace para mitigar percances meteorológicos (tornados, inundaciones), incendios o graves errores humanos. Actualmente todos los grandes proyectos de preservación digital incluyen la replicación remota; algunos ejemplos son: Library of Congress, US National Archives and Record Administration, Bibliothèque Nationale de France, Archives Nationales de France, Internet Archive, University of California Berkeley y New York Public Library.

En el caso de instituciones y empresas pequeñas y medianas la replicación de todo el sistema informático resulta inviable por sus altos costes. Aun en este caso es imprescindible que las copias de seguridad de los datos y del software se guarden de forma distribuida, con al menos una copia alejada de las instalaciones principales. En casos graves tiene poco valor haber guardado una copia de los datos del ordenador en el cajón del propio escritorio, como hace mucho personal administrativo, pues, por ejemplo, un incendio o un derrumbe del edificio van a afectar por igual al disco duro del ordenador y a la copia guardada.

Algunas buenas prácticas para asegurar la disponibilidad de estas copias distribuidas son: contratar a una empresa especializada para que cada semana recoja y guarde de forma alejada una copia de seguridad, intercambiar copias de seguridad entre instituciones similares (por ejemplo entre dos universidades) o mantener una de las copias en sistemas de almacenamiento en la nube como Amazon Glacier (http://aws.amazon.com/es/glacier/)

Migración de formatos

Los formatos de los ficheros pueden convertirse en obsoletos con el paso del tiempo, por lo que es necesario migrar los contenidos a un formato más actual para continuar teniendo acceso a los mismos. En otras ocasiones la migración se realiza de forma voluntaria porque se desean obtener las ventajas de otro formato. Algunas de las razones para migrar de formato son:

- Han aparecido versiones más modernas; por ejemplo, pasamos de Microsoft Word 95 a Microsoft Word 2010.
- Deseamos pasar a un formato no propietario, estándar: de documentos en Microsoft Word 95 a LibreOffice. Este es un movimiento que se está dando en muchas administraciones públicas.
- Deseamos pasar a formatos no compilados o de código abierto: de bases de datos Microsoft Acces a MySQL.
- Deseamos usar los estándares industriales *de facto*; muchas empresas prefieren concentrar sus datos en formatos de Microsoft Office («así mis problemas serán los de la mayoría de la humanidad»).
- Abandonamos formatos de fabricantes ya no existentes: dejamos el gestor de base de datos dBase, de Ashton-Tate.

- Deja de funcionar el viejo hardware, tenemos que pasar a otro más nuevo pero que no admite los formatos antiguos: abandonamos el ordenador Sinclair ZX Spectrum.
- Requerimientos legales, como haber caducado las licencias de uso o, por el contrario, tener que pagar licencias a partir de determinado momento. Es el caso de la migración de los gráficos GIF (formato bajo *copyright* de una empresa privada) a PNG (formato libre) y también, por la misma razón, del paso del formato de sonido comprimido MP3 a Ogg Vorbis.
- Se unifican los formatos de los ficheros para facilitar su gestión conjunta. En el caso de archivos que reciben transferencias directamente de los productores, como las oficinas de una administración pública o los investigadores de una universidad, es normal que acumulen documentos parecidos pero grabados con formatos distintos (por ejemplo, distintos formatos de procesador de textos o de hoja de cálculo). Esta heterogeneidad va en contra de una gestión eficiente, que es más fácil con el manejo de un número pequeño de formatos de los que se conocen sus características. Para solucionar estos casos, los archivos de preservación normalizan los ficheros que reciben, es decir, los migran de sus formatos originales a unos pocos formatos que han seleccionado como idóneos para su conservación.

Vemos que las causas para emprender una migración de datos son variadas, por ello es importante que las organizaciones realicen un buen análisis de su situación presente y planifiquen cuáles son las propiedades que más van a valorar en el futuro: el soporte de un fabricante, la estandarización del formato, el libre acceso al código fuente, su gratuidad, el nivel de aceptación en el mercado, etc. Por tanto, ante un mismo conjunto de ficheros, dos entidades pueden apostar por formatos distintos y las dos decisiones pueden ser igualmente acertadas. A partir del análisis previo se habrán de planificar las migraciones vigilando cuatro puntos:

- En cada migración se corre el riesgo de pérdida de funcionalidades o de propiedades significativas de los documentos, pues ningún formato es realmente equivalente a otro.
- En cada migración se corre un riesgo de pérdida de datos y de producirse errores que se acumulan a los generados en migraciones anteriores; por ello es importante reducir el número de migraciones o, lo que es lo mismo, alargar el tiempo de vida de los formatos; cuantos menos cambios, mejor.

- Una migración supone un gran esfuerzo técnico y económico por los recursos humanos e informáticos que consume. En muchos casos ejecutar una migración en un formato de datos impone cambios en el software que lo visualiza y gestiona.
- Los tiempos necesarios para reprocesar los ficheros pueden ser enormes si nos encontramos ante grandes volúmenes de datos. Ello obliga a planificar con gran antelación los recursos necesarios y el momento en que se va a ejecutar la migración.

No debemos olvidar que después de una migración no se ha terminado el problema, simplemente empieza otro ciclo de vida de los ficheros hasta que llegue el momento de la próxima migración. Cualquiera que sea nuestra política de migraciones, esta tendrá que estar documentada y las acciones que a lo largo del tiempo se vayan ejecutado sobre un fichero o conjunto de ficheros determinado habrán de quedar registradas, mediante metadatos PREMIS o de otro tipo. Esta información nos permitirá seguir la trazabilidad de los ficheros a lo largo del tiempo y con ella justificar sus cambios de formato y asegurar ante terceros la autenticidad de su contenido. Tener documentadas las migraciones también posibilita descubrir en un futuro cuál es la causa de un determinado error de formato y gracias a ello, quizás, paliarlo.

Comprobamos que la migración de formatos es en muchos casos una actividad obligada pero peligrosa por los errores que puede generar (aunque seguramente puede ser mucho peor no hacer nada) y costosa. Por ello las políticas de migración están en el centro de las investigaciones en preservación digital, con el fin de encontrar alternativas a las mismas o, como mínimo, minimizar su impacto. Veamos a continuación dos reflexiones que se están haciendo sobre el tema.

La primera plantea la siguiente pregunta: si el formato original de los ficheros es migrado a un nuevo formato para su mejor preservación, ¿qué pasa con el fichero original? ¿Lo descartamos (borramos) porque ya tenemos una copia mejor o también lo conservamos como testimonio auténtico que fue? La respuesta no es fácil ni es universal: depende. Si la migración ha sido efectuada en condiciones técnicas controladas, se dispondrá de mecanismos para asegurar y, si es necesario, certificar que el nuevo fichero mantiene los atributos de autenticidad e integridad del fichero original, con lo que este podría ser destruido. Aun así, distintos archivos históricos nacionales están

preservando ambas copias, pero esta es una política que a medio y largo plazo no se podrán permitir aplicar a la totalidad de sus fondos, pues lleva, como mínimo, a duplicar las necesidades de sistemas de almacenamiento y este es un gasto inasumible por cualquier país.

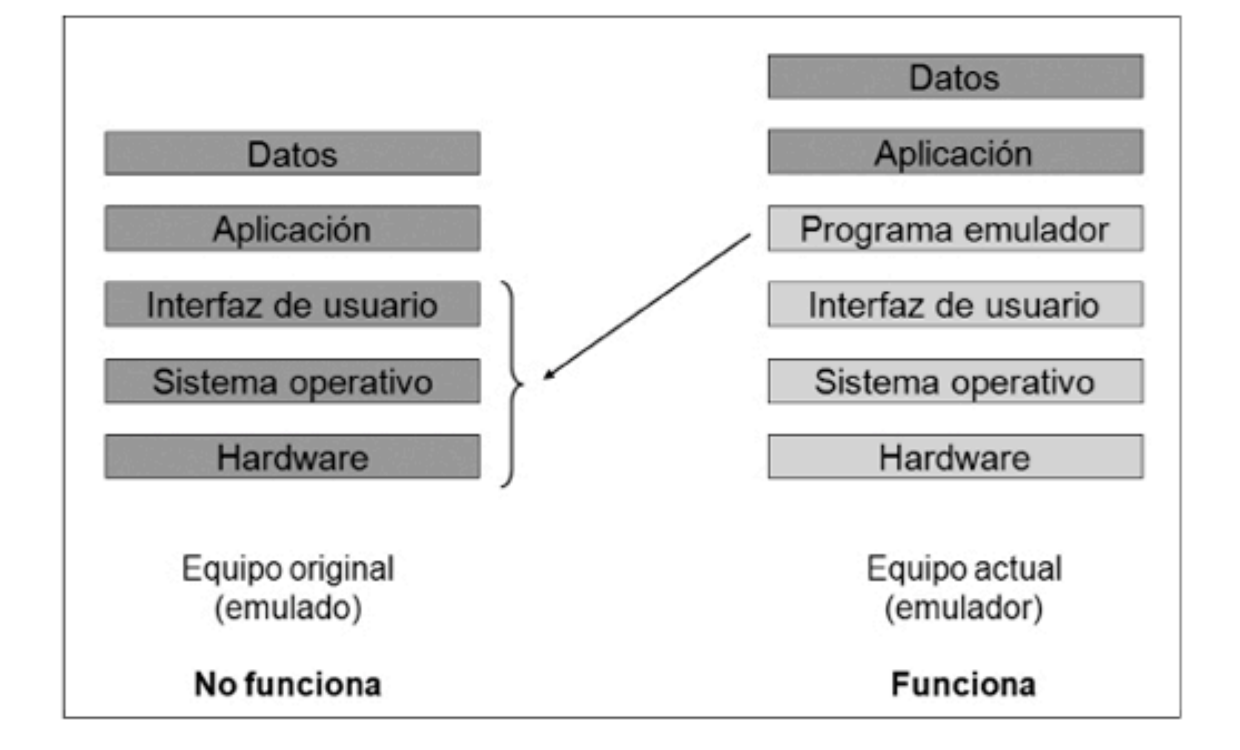
La segunda reflexión versa sobre el hecho de que, como ya ha ocurrido con los documentos analógicos (sobre todo en papel), muchos documentos de archivo o de biblioteca nunca llegarán a ser consultados por los usuarios del futuro o solo lo serán de forma esporádica al cabo de mucho tiempo. Si es así, puede ser un despilfarro migrar ficheros que quizás nunca se van a usar; ante esto algunos expertos proponen realizar, si así se acordó, la normalización de formatos en el momento de la ingestión y no ejecutar nuevas migraciones hasta el mismo momento en que un usuario necesite usar el fichero.

Emulación

Distintos investigadores, entre los que destaca Jeff Rothenberg, han indicado que la migración de formatos no es un método sostenible a largo plazo y aplicable a grandes volúmenes de datos, debido a sus costes y a los riesgos técnicos inherentes que ya hemos mencionado. Para estos investigadores la solución más eficiente es la emulación: crear un programa (el emulador) que simule el funcionamiento de hardware o de software del pasado (los emulados) y así permita acceder a los datos, que aún se mantienen guardados en sus formatos originales, sin migración.

Existen dos variantes de la emulación: la emulación de hardware y la emulación de software. En la primera disponemos de software y datos que solo funcionan en un hardware antiguo que ya no funciona o no existe; sería el caso de programas que solo pueden funcionar en un ordenador PC-AT de principios de los años ochenta. En esta situación el programa emulador simula el funcionamiento del ordenador antiguo. En la segunda variante, la emulación de software, el software original ya no existe o no funciona con los ordenadores actuales, entonces un software emulador simula su funcionamiento, permitiendo acceder a los datos archivados.

Ilustración 3. Esquema de funcionamiento de un emulador de hardware



Algunos ejemplos de usos actuales de la emulación de hardware son:

- Los emuladores que proporcionan los fabricantes de teléfonos móviles a los desarrolladores de programas para los mismos. Estos emuladores permiten que desde un ordenador personal se pueda realizar cómodamente la programación y la prueba de los programas que se están desarrollando, sin depender de la pequeña pantalla y del teclado del teléfono a los que van destinados.
- Los emuladores que simulan desde un PC el funcionamiento de las antiguas máquinas recreativas o de las actuales consolas y de esta forma permiten jugar con multitud de juegos de las respectivas categorías. Hay que advertir que mientras el emulador en sí es un programa legal, no lo es la incorporación de los juegos que, salvo raras excepciones, han sido previamente pirateados y ensamblados de nuevo de forma ilegal sin el permiso previo de sus propietarios. MAME (http://mamedev.org/) y MESS (http://www.mess.org/) son los ejemplos más destacados de este tipo de software emulador. MAME es capaz de emular la práctica totalidad de máquinas recreativas y de consolas de juegos; MESS, a su vez, es una transformación de MAME especializada en la emulación de microordenadores de todas las épocas y fabricantes.

En diversas instituciones se están probando emuladores de distintos tipos de ordenadores de los años setenta y ochenta, a fin de disponer de unos procedimientos que permitan la recuperación de los contenidos creados en

esos años. En esta línea destaca el proyecto Dioscuri (http://dioscuri.sourceforge.net/), un emulador de la arquitectura x86 de los ordenadores PC de la década de 1980, que ha sido creado por la biblioteca nacional de los Países Bajos dentro de los proyectos europeos Planets y Keep. Un segundo nivel en el que ahora se está trabajando es la emulación a distancia (el programa emulador se ejecuta a distancia por lo que ya no es necesario tenerlo instalado en un entorno local) y los metaemuladores, como Emulation Framework, creado por el proyecto europeo Keep (http://emuframework.sourceforge.net/).

Los emuladores de software no están de momento desarrollados, pues se ha dado prioridad a la emulación de hardware antiguo, al tratarse de soluciones de aplicación más general. Por otro lado, en muchas ocasiones los formatos antiguos pueden recuperarse con programas actuales distintos de aquellos con los que fueron creados, con lo que no es imprescindible contar con el programa original o emularlo. Un ejemplo es la hoja de cálculo Lotus 123, la más utilizada en la década de 1980, un formato que hoy en día aún se puede abrir desde programas como Microsoft Excel y LibreOffice Calc. Esta solución no es de aplicación universal, pues los textos escritos en la misma época con WordStar ahora no se pueden abrir con ningún programa habitual. Otra solución que ha aparecido últimamente es la de programas que actúan como multivisores y permiten visualizar pero no editar los contenidos de formatos obsoletos; un buen ejemplo de esta categoría es el programa Quick View Plus (http://www.avantstar.com).

Se debe recordar que ningún emulador es perfecto, es decir, ningún emulador actúa igual que el elemento emulado. Hoy, por ejemplo, podemos emular los primeros ordenadores personales con una pequeña pantalla de fósforo verde y el sonido «bip-bip» procedente de su altavoz interno, pero lo hacemos en ordenadores con grandes pantallas de alta resolución en color y potentes tarjetas de sonido. Se podrá decir que las condiciones técnicas actuales son mejores, pero en ningún caso idénticas a las originales, pregúntenlo si no a los aficionados a los primeros juegos electrónicos, que desean volver a matar con rayos verdes a marcianos que solo sueltan un monocorde «bip-bip» al caminar.

Análisis forense digital

Las técnicas y metodologías de preservación digital prevén que todos los ficheros serán correctamente tratados y podrán ser interpretados en el futuro, pero esto no sucederá en todos los casos. A pesar de nuestras buenas intenciones habrá ficheros que llegarán al futuro sin haber sido tratados. Este problema actualmente ya empieza a detectarse; fijémonos en los casos en que un archivo o una biblioteca recibe el legado documental de un escritor, un político u otra persona eminente como resultado de una previsión testamentaria o de una donación en su vejez. Si este tipo de fondos antes estaba compuesto básicamente por documentos en papel, actualmente empiezan a incorporar disquetes o incluso discos duros enteros. Estos soportes informáticos contienen ficheros con las versiones originales de algunas de sus obras e incluso pueden albergar obras inéditas, además de correspondencia privada (en forma de correo electrónico o con otros formatos) y otra documentación de interés. A menudo se trata de sistemas de almacenamiento que actualmente ya no pueden ser leídos de forma directa (como disquetes de 5 ¼", discos Iomega Zip...) que, a su vez, guardan ficheros grabados en formatos antiguos (WordStar, Lotus 123...). Los centros receptores han de ser capaces de leer estos soportes, recuperar sus contenidos, interpretarlos, documentarlos y transferirlos a medios y formatos actuales, legibles por los investigadores actuales.

Las bibliotecas y archivos nacionales, así como bibliotecas universitarias importantes, se están dotando a nivel internacional de los medios técnicos para poder tratar este tipo de donaciones de materiales informáticos. Las técnicas aplicadas de recuperación a veces se conocen con el nombre de arqueología digital, pero desde un punto de vista riguroso se trata de un subconjunto de las técnicas de análisis forense digital, normalmente usadas por las fuerzas policiales y por investigadores privados para investigar delitos realizados en el ámbito informático, como pueden ser: piratería informática, espionaje industrial, sabotaje, ciberterrorismo, accesos prohibidos, pornografía infantil, etc.

En análisis forense aplicado a arqueología digital básicamente nos encontramos con tres tipos de problemas: ser capaces de leer el soporte

hardware, conseguir navegar dentro de los contenidos para localizar los que nos sean de interés y saber leer y, si es necesario, migrar los formatos obsoletos.

Nos podemos encontrar con dos tipos de soportes: removibles (disquetes) y fijos (discos duros). En donaciones es habitual recibir disquetes en formatos antiguos (5 ¼", 3 ½", Iomega Zip...), sin que dispongamos de los lectores necesarios para su lectura. La posibilidad de conservar máquinas antiguas con las que mantener la capacidad de leer medios antiguos parece la solución más fácil pero es inviable a medio plazo, pues estas máquinas llega un momento en que acaban fallando debido a su antigüedad y a la evidencia de que resulta cada vez más difícil proceder a su reparación. Por otro lado, en estas máquinas antiguas, que funcionan con sistemas operativos también obsoletos como CP/M, MS-DOS o Windows 95, no es posible instalar software moderno que permita la migración de los ficheros que buscamos y a menudo tampoco permiten conectar medios modernos de almacenamiento con conexiones USB o SATA para el traspaso de los ficheros. En el mercado especializado se venden placas controladoras y lectores para la mayoría de formatos, pero su conexión a un ordenador corriente no es trivial; por ejemplo, las unidades lectoras de 5 ¼" ya no se pueden conectar a la placa base de un ordenador actual, que no dispone ni de los circuitos impresos ni del tipo de conexión adecuada.

La alternativa está en la utilización de hardware especializado que permite conectar y acceder a medios antiguos desde ordenadores y sistemas operativos actuales. Son ejemplos de este tipo de hardware las máquinas forenses FRED (http://www.digitalintelligence.com/products/fred/) y las controladoras de disquetes Kryoflux (http://www.kryoflux.com/). Se recomienda que estos sistemas dispongan de funciones de bloqueo de escritura (*write blocked*) con los que se puedan mantener inalterados los originales y así asegurar la autenticidad de los ficheros que se obtengan. También es necesario disponer del software que sepa leer el tipo de grabación de los datos, pues existen grabaciones de alta y baja densidad, para funcionar en entornos Amiga, MS-DOS o Mac, por ejemplo. La conexión de discos duros (IDE, SCSI, SATA) es mucho más fácil, pues los cambios técnicos han sido menores y se venden dispositivos conectores de bajo coste

El segundo reto es conseguir navegar en el interior del medio de almacenamiento sin alterar su contenido, pues ello podría ir en contra del aseguramiento de su autenticidad. En todos los casos será necesario usar

soluciones de software o de hardware que impidan la grabación accidental del medio al que accedemos; en esta línea son muy útiles los conectores o los duplicadores de discos con tecnología *write blocked*, como los de los fabricantes Tableau (http://www.tableau.com/) y WiebeTech (http://www.wiebetech.com/). De esta manera podremos navegar de forma segura dentro del disco o, aún mejor, generar una imagen, una copia idéntica del mismo con la que trabajar.

Para el análisis del contenido se deben usar programas especializados, entre los que destacan los comerciales Encase, de Guidance Software, Forensic Toolkit (FTK), de AccessData, X-Ways y los programas libres Autopsy Sleuth Kit, CAINE y Helix3. Estos programas nos permitirán descubrir los ficheros que buscamos, extraer sus metadatos técnicos y comprobar su autenticidad.

- Encase: http://www.guidancesoftware.com/encase-forensic.htm
- Forensic Toolkit (FTK): http://www.accessdata.com/products/digital-forensics/ftk
- X-Ways: http://www.x-ways.net/
- Autopsy Sleuth Kit: http://www.sleuthkit.org/
- CAINE: http://www.caine-live.net/
- Helix3: http://www.e-fense.com/helix3pro.php

La tercera etapa consiste en migrar los ficheros recuperados a un formato actual y almacenarlos de forma segura en el repositorio de la institución. La migración completa de formatos antiguos no siempre es posible, aunque sí ha de ser viable la recuperación del texto de los documentos.

Algunas de las iniciativas ligadas al movimiento a favor de los datos abiertos (*open data*) han empezado a generar una ampliación en los usos posibles de las técnicas de análisis forense. Distintas universidades y centros de investigación están creando repositorios institucionales de datos abiertos, en los que se depositan y se conservan los datos primarios que se recogieron o generaron en el transcurso de un proyecto de investigación determinado. Estos datos primarios pueden estar fragmentados en multitud de ficheros y no siempre en formatos de uso común, por lo que su ingestión, catalogación, conservación y puesta a disposición de los usuarios puede resultar una tarea demasiado voluminosa y complicada técnicamente para los procedimientos habituales de las bibliotecas. Por ello, algunas de ellas están tomando la determinación de

gestionar estos contenidos no fichero a fichero sino en bloque, a nivel de unidades de almacenamiento (por ejemplo, todo un disco externo USB). En esta situación las herramientas de análisis forense permiten realizar una copia en imagen de la unidad de almacenamiento, asegurar su autenticidad e integridad y facilitar, más tarde, su consulta por medio de máquinas virtuales o mediante una navegación simulada por su árbol de ficheros.

HERRAMIENTAS Y ESTÁNDARES DISPONIBLES

Herramientas

Cada vez hay disponibles más herramientas de software que facilitan el trabajo de preservación digital. A continuación vamos a mostrar en qué consisten algunas de estas herramientas.

De comprobación de formatos

Para que un sistema de preservación sea realmente responsable de la custodia y preservación de unos ficheros es imprescindible que conozca en qué formatos están codificados y cuáles son los peligros de conservación de los mismos (básicamente si aún son de uso común y se dispone de software para su utilización). Por ello, como ya vimos anteriormente, el modelo OAIS prevé que en la fase de Ingesta el sistema compruebe y registre el formato de cada fichero ingresado y que este metadato sea traspasado a Gestión de datos.

La experiencia demuestra que existen muchos ficheros con la extensión errónea o que un mismo contenido se puede encontrar con distintas extensiones. Por otro lado, las primeras instituciones que empezaron a actuar en el ámbito de la preservación digital comprobaron que no se disponía de un registro mundial de todos los formatos existentes, de cuáles eran sus características técnicas y de cómo identificarlos de forma fehaciente. Para remediarlo, tanto la Library of Congress como los archivos nacionales del Reino Unido, entre otros, empezaron a crear registros públicos y fiables de los formatos de ficheros más habituales.

Con el paso del tiempo el registro de los archivos nacionales británicos se ha consolidado como el más completo y actualmente es usado por muchas organizaciones de distintos países; este registro se llama Pronom (http://www.nationalarchives.gov.uk/pronom) y es de consulta pública. Para facilitar su uso, la misma organización creó el software gratuito DROID para su

consulta automatizada. DROID es capaz de identificar los formatos almacenados en un sistema externo (un disco duro, una memoria USB) a partir de las indicaciones técnicas contenidas en Pronom. El número de formatos reconocidos por Pronom está en continuo crecimiento gracias a las aportaciones técnicas de multitud de colaboradores.

De forma paralela, en Estados Unidos la Universidad de Harvard ha creado otro programa, JHove, que no solo identifica los formatos de los ficheros sino que en algunos casos también indica si están bien formados respecto a la norma técnica original y es capaz de extraer un gran volumen de metadatos técnicos.

Droid y JHove son soluciones robustas y fiables a nivel técnico pero a veces son criticadas por no tener un funcionamiento lo suficientemente rápido como para procesar grandes volúmenes de ficheros o por no reconocer un número mayor de formatos. Otras aplicaciones resuelven estos problemas, aunque sea a costa de aplicar algoritmos de identificación no tan seguros, como FIDO, TrID, File Investigator Engine y Outside-In File ID, las dos últimas a nivel comercial.

- DROID: http://sourceforge.net/projects/droid/
- JHove: http://sourceforge.net/projects/jhove/
- FIDO: http://www.openplanetsfoundation.org/software/fido
- TrID: http://mark0.net/
- File Investigator Engine: http://www.forensicinnovations.com/
- Outside-In File ID: http://www.oracle.com/us/technologies/embedded/025613.htm

Ante esta dispersión de esfuerzos, en los últimos años distintas instituciones han mantenido contactos para intentar agregar algunas de las iniciativas que hemos mencionado y crear un registro mundial de formatos. Fruto de estos contactos a mediados de 2012 se abrió el Unified Digital Formats Registry UDFR; (http://udfr.org/), creado por la Universidad de California y con el apoyo de distintas bibliotecas y archivos nacionales, pero de momento no ha conseguido obtener la suficiente financiación como para sostener sus actividades. Por ello Pronom continúa siendo la mejor opción para el reconocimiento de formatos y su identificador PUID uno de los más usados.

De comprobación de la integridad (*checksum*)

Comprobar que los bytes que forman un fichero no se han visto alterados desde un momento previo es una de las técnicas primordiales para asegurar que ese fichero no ha sido manipulado intencionadamente o ha cambiado accidentalmente (por ejemplo, por un error de grabación), con lo que aseguramos su integridad. Que un fichero sea íntegro no asegura al mismo tiempo la autenticidad de su contenido (este ya podría ser falso desde el momento de su creación), pero sí nos permite eliminar la incertidumbre de saber si es exactamente el mismo que fue creado o no.

El método más habitualmente usado para comprobar la integridad de ficheros es el cálculo de la suma de verificación (o *checksum*), un tipo de función *hash* que mediante la aplicación de un algoritmo matemático calcula la equivalencia del conjunto de bits de un fichero a un valor discreto. Las funciones *hash* que actualmente se están usando son MD5, SHA-1 y SHA-256. No se ha de confundir la suma de verificación de un documento con su firma digital generada con certificado electrónico: la primera es única y distinta para cada documento, mientras que la segunda es única para cada autor. En aplicaciones de administración electrónica o de comercio electrónico es habitual trabajar con una combinación de ambos sistemas, añadiendo al documento una firma electrónica que incluye una suma de verificación, así se identifica de forma inequívoca quién ha creado el documento y además permite comprobar la integridad del mismo

Ejemplo 1. Suma de verificación SHA-1

```
23D01C3A10B7792A5CDA34564097C8E5358ACE75
```

Ejemplo 2. Código parcial de una firma electrónica generada mediante un certificado digital de la FNMT de España

20 82 01 22 30 0D 06 09 2A 86 48 86 F7 0D 01 01 01 05 00 03 82 01 0F 00 30 82 01 0A 02 82 01 01
00 B4 60 37 B6 1F BE FB DA FE 3F CE 16 DD F1 62 A9 D4 55 B8 FC 1D 5C 52 2C D6 F7 D5 32 EB A4 FB
11 57 31 BC 61 25 8D 6A 6D 75 74 C0 C8 06 31 E1 CF A2 8E B8 96 75 68 F4 24 27 65 6E B9 F1 0E 69 50
91 70 2C 1A A6 CD 69 A9 31 8D 61 36 4E 8A 22 96 CB 56 4D 3F 79 34 71 F3 5A D3 48 A8 64 24 15 7F
7D 66 11 8A 46 F8 0B 8F 0C 78 99 74 66 92 40 2C 50 B6 AF 7F 2F 83 CB 30 2D 40 32 69 4E 27 BA A2 98
63 64 1F 3D B9 2A B4 4F 55 FD 78 F9 D0 8A 14 EF 88 1D 97 C0 0F B7 2E 2C 51 A9 BE 46 39 CD C3 75
76 58 A9 E8 2C 13 20 0B 86 49 A3 CA C8 12 2E 3A A5 2C C0 C5 56 B9 09 E6 C9 E1 F5 A9 5E 1F 57 3E 9A
4D 9F 28 AD E4 75 55 AB 4B 2E A5 C5 0A 23 3C 36 0A EE 9C 7F 42 A5 37 80 92 05 50 BA DF D5 6D E0
E7 CB CE C6 B5 AD 80 B3 67 3A 9A D6 3A D6 28 72 1E D6 73 46 0E 64 7D D9 43 09 1D 69 00 E3 02 03
01 00 01

En la preservación digital lo correcto es que los sistemas de copiado de ficheros integren la generación y la comprobación de sumas de verificación, y que también sean usadas en las transferencias de ficheros con el fin de asegurar que los ficheros recibidos son los mismos que los enviados por el productor o la institución colaboradora. El cálculo de los valores *hash*, la generación de ficheros con listados de estos valores, así como la comprobación de su validez, son funciones integradas en sistemas profesionales de copiado y también están soportadas por numerosos programas de código libre.

De transferencia de ficheros

Muchas instituciones y empresas de tamaño mediano y pequeño no tienen ni tendrán en el futuro la suficiente capacidad técnica, económica y/u organizativa para ejecutar directamente políticas de preservación a largo plazo. En otros casos, siguiendo normativas administrativas o legales, determinados fondos deberán transferirse a instituciones o empresas especializadas en la preservación, como son los archivos y las bibliotecas nacionales o las empresas especializadas en la custodia de documentación electrónica, como Iron Mountain (http://www.ironmountain.com/). Por todo ello, la transferencia de fondos digitales entre entes distintos será una práctica habitual en el futuro y ya empieza a serlo en algunos países.

Este intercambio se ha de realizar de forma eficiente y segura, cosa que a nivel técnico implica la normalización de los formatos de intercambio y de los protocolos de transmisión o transporte. Actualmente cuando dos personas desean intercambiarse ficheros, porque por ejemplo están realizando un trabajo en equipo, es habitual que el emisor empaquete los ficheros en un único

fichero comprimido zip o similar, que presenta diversas ventajas: disminuye su volumen, evita el peligro de olvidarse algún fichero, mantiene si es necesario la jerarquía de directorios y permite una fácil comprobación de la integridad de la transmisión. De forma parecida a este ejemplo, se están implementando procedimientos para transmitir de forma automatizada grandes volúmenes de ficheros, de forma que se pueda asegurar su integridad e informar de la estructura del contenido. Son ejemplos de estos procedimientos técnicos de transmisión de ficheros:

- NISO Journal Article Tag Suite (JATS): http://jats.nlm.nih.gov/. Antes NLM DTD, es un esquema XML creado por la National Library of Medicine de Estados Unidos para codificar la estructura de las revistas científicas electrónicas y transmitirlas entre los editores y las bibliotecas cliente.
- Bag-It: https://wiki.ucop.edu/display/Curation/BagIt. Es un programa desarrollado por la Library of Congress y la California Digital Library para transmitir grandes volúmenes de ficheros a centros de preservación digital. Se está usando bastante en las transferencias entre grandes instituciones de Estados Unidos y para generar paquetes SIP a ingestar en un sistema OAIS.

Estos estándares de empaquetado facilitan la transmisión y también la ingestión automatizada de los ficheros en el sistema informático receptor, por lo que se están extendiendo como rutinas para la importación y la exportación de datos entre sistemas informáticos diferentes.

Estándares de metadatos

Junto con los documentos también se deben presentar los metadatos que faciliten su gestión por el sistema de preservación así como los necesarios para facilitar su uso por los posibles usuarios. Estos últimos variarán según la naturaleza de los documentos y del sector donde son usados. En el caso de colecciones alojadas en archivos, bibliotecas y museos son de uso común metadatos orientados a la descripción intelectual del contenido (DC, MARCXML, MODS, EAD, PBCore...), a su descripción técnica (NISO MIX, EXIF, XMP...) y al marcaje de sus condiciones de propiedad intelectual (METSRights...).

En la primera categoría de metadatos que hemos mencionado al principio, aquellos propios de los sistemas de preservación, se encuentran METS y PREMIS, que a continuación se describen con detalle.

METS

METS (http://www.loc.gov/standards/mets/), o Metadata Encoding and Transmission Standard, es un esquema XML para codificar metadatos relacionados con la preservación de objetos digitales, mantenido por la Library of Congress. METS explica cómo codificar metadatos descriptivos (título, autor, editor...), de propiedad intelectual (procedencia y derechos de uso de los contenidos), técnicos (profundidad de color, resolución, frecuencia de muestreo, tamaño en bytes, valor de verificación *hash*, localizador URI...) y estructurales (qué ficheros representan este objeto digital y cómo se relacionan entre ellos).

Es extensible y permite que se puedan codificar los distintos tipos de metadatos según estándares XML externos, como pueden ser DC, MARCXML y PREMIS. Pero su punto fuerte son las secciones de ficheros y estructural; en ellas los objetos digitales compuestos, formados por diversos ficheros (como es el caso de libros con las páginas escaneadas por separado), tienen marcada su jerarquía (cuáles actúan como índice o como punto de entrada) y ordenación (en qué orden se han de presentar o usar). METS también mantiene la relación entre distintas manifestaciones conservadas de una misma obra (por ejemplo, un libro en PDF, TIFF, JPEG, EPUB y TXT) y la función que desempeña cada una de ellas (por ejemplo, miniatura de previsualización, versión de consulta por pantalla, versión de archivo, versión de descarga y texto OCR para búsquedas).

En el Ejemplo 3 se muestran dos secciones del esquema METS de un hipotético documento que fue escaneado y se conserva dividido en dos ficheros. Esta institución conserva el documento en cuatro versiones digitales distintas, cada una de ellas formada por dos ficheros: ficheros TIFF máster, ficheros JPEG para la consulta por pantalla, ficheros JPEG para la previsualización como miniaturas y ficheros TXT con el texto obtenido por OCR.

Se está usando dentro de la Gestión de datos de sistemas OAIS y también como sistema de transmisión (importación/exportación) de contenidos digitales entre aplicaciones o entre instituciones. Por ello ya se ha convertido en

un requerimiento casi obligatorio de cualquier software gestor de repositorios y de los gestores documentales, pues METS facilita la integración de colecciones generadas externamente o la migración entre sistemas informáticos distintos.

Una mención especial se merece METS-ALTO (http://www.loc.gov/standards/alto/), una extensión de METS que permite describir la maquetación y el contenido de prensa histórica que ha sido digitalizada y su texto extraído por OCR.

Ejemplo 3. Secciones de ficheros y estructural de un esquema METS. Se trata de un objeto compuesto por dos ficheros, que se presenta en cuatro versiones distintas

```
<fileSec>
	<fileGrp USE="master">
		<file ID="id1" MIMETYPE="image/tiff" SEQ="1">
			<FLocat xlinkhref="http://servidor.com/masters/fichero0010.tiff"
			LOCTYPE="URL"/>
		</file>
		<file ID="id2" MIMETYPE="image/tiff" SEQ="2">
			<FLocat xlinkhref=" http://servidor.com/masters/fichero0011.tiff "
			LOCTYPE="URL"/>
		</file>
	</fileGrp>
	<fileGrp USE="consulta">
		<file ID="id3" MIMETYPE="image/jpg" SEQ="1">
			<FLocat xlinkhref="http://servidor.com/consulta/fichero0010.jpg"
			LOCTYPE="URL"/>
		</file>
		<file ID="id4" MIMETYPE="image/jpg" SEQ="2">
			<FLocat xlinkhref=" http://servidor.com/consulta/fichero0011.jpg
			"LOCTYPE="URL"/>
		</file>
	</fileGrp>
	<fileGrp USE="miniatura">
		<file ID="id5" MIMETYPE="image/jpg" SEQ="1">
			<FLocat xlinkhref="http://servidor.com/miniaturas/fichero0010.jpg"
			LOCTYPE="URL"/>
		</file>
		<file ID="id6" MIMETYPE="image/jpg" SEQ="2">
			<FLocat xlinkhref=" http://servidor.com/miniaturas/fichero0011.jpg "
			LOCTYPE="URL"/>
		</file>
	</fileGrp>
	<fileGrp USE="OCR">
		<file ID="id7" MIMETYPE="text/plain" SEQ="1">
			<FLocat xlinkhref="http://servidor.com/txt/fichero0010.txt" LOCTYPE="URL"/>
		</file>
		<file ID="id8" MIMETYPE="text/plain " SEQ="2">
			<FLocat xlinkhref="http://servidor.com/txt/fichero0011.txt " LOCTYPE="URL"/>
		</file>
	</fileGrp>
</fileSec>
<structMap TYPE="presentacion_fisica">
	<div ORDER="1" TYPE="text" LABEL="Paginacion">
		<div ORDER="1" TYPE="pagina" LABEL="[p. 1]">
			<fptr FILEID="id1"/>
			<fptr FILEID="id3"/>
			<fptr FILEID="id5"/>
			<fptr FILEID="id7"/>
		</div>
		<div ORDER="2" TYPE="pagina" LABEL="[p. 2]">
			<fptr FILEID="id2"/>
			<fptr FILEID="id4"/>
			<fptr FILEID="id6"/>
			<fptr FILEID="id8"/>
		</div>
	</div>
</structMap>
```

PREMIS

PREMIS (http://www.loc.gov/standards/premis/), mantenido por la Library of Congress, es un esquema XML que implementa un catálogo de metadatos de preservación. PREMIS se centra en documentar las propiedades técnicas, los agentes, los derechos y los eventos implicados en la preservación de un objeto digital a lo largo de su vida. Su objetivo es dejar constancia documentada, de una forma estandarizada y por tanto automatizable, de aquellos temas que puedan afectar a la conservación de un objeto digital. Dentro de un modelo OAIS, PREMIS forma parte del núcleo de metadatos de la Gestión de datos.

Algunos ejemplos de eventos que se documentan con PREMIS son: migraciones de formatos, cambios en la política de *hash*, controles de antivirus aplicados, controles de integridad, cambios en los derechos de propiedad intelectual, etc. Documentar estos eventos con PREMIS presenta la ventaja de facilitar la ejecución posterior de acciones automatizadas de mantenimiento sobre grandes conjuntos de ficheros que presentan las mismas propiedades (por ejemplo: reformatear todos los ficheros del formato A que el año 2008 se migraron equivocadamente desde el formato B con el software defectuoso C). PREMIS se puede entender, por tanto, como una recopilación preventiva de información de ficheros en previsión de posibles actuaciones posteriores, de forma similar a como la historia clínica de los pacientes de un centro médico recoge su historial de dolencias y tratamientos para facilitar posibles acciones futuras.

En muchas instituciones PREMIS se implementa como una extensión de METS, pues se prefiere integrar todos los metadatos en un solo conjunto en el que METS ordena las distintas categorías de metadatos. Se ha de advertir que aunque el catálogo de metadatos PREMIS es muy amplio, en la mayoría de casos es necesario adaptar su uso a las pautas de trabajo de una institución concreta. Por esta razón, en el momento actual es necesario efectuar un análisis de los requerimientos de cada situación e incluso desarrollar *scripts* y conectores que faciliten la gestión automatizada de los metadatos PREMIS.

En el Ejemplo 4 se muestra un registro PREMIS 2.1 completo. Corresponde a un fichero JPEG del que se da el valor de su *hash* MD5, su peso en bytes, su identificación como formato MIME, su identificación según el directorio de

formatos Pronom y su URI; también incluye un evento completo: su validación con el programa JHove y el resultado de la misma.

Ejemplo 4. Metadatos PREMIS. Se trata de los metadatos de un fichero JPEG, del que se registran distintos valores de entrada y de almacenamiento, así como los datos de un evento de validación


```xml
<?xml version="1.0" encoding="utf-8"?>
<premis xmlns="info:lc/xmlns/premis-v2" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="info:lc/
xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2-1.xsd" version="2.1">
	<object xsi:type="premis:file">
		<objectIdentifier>
			<objectIdentifierType>FILE</objectIdentifierType>
			<objectIdentifierValue>001</objectIdentifierValue>
		</objectIdentifier>
		<objectCharacteristics>
			<compositionLevel>0</compositionLevel>
			<fixity>
				<messageDigestAlgorithm>MD5</messageDigestAlgorithm>
				<messageDigest>dd21d99a468f3bb52a136ef5beef5034</messageDi-
				gest>
			</fixity>
			<size>200001</size>
			<format>
				<formatDesignation>
					<formatName>image/jpeg</formatName>
				</formatDesignation>
				<formatRegistry>
					<formatRegistryName>PRONOM</formatRegistryName>
					<formatRegistryKey>fmt/43</formatRegistryKey>
				</formatRegistry>
			</format>
		</objectCharacteristics>
		<storage>
			<contentLocation>
				<contentLocationType>URL</contentLocationType>
				<contentLocationValue>http://servidor.com/consulta/fichero0011.
				jpg</contentLocationValue>
			</contentLocation>
		</storage>
	</object>
	<event>
		<eventIdentifier>
			<eventIdentifierType>LocalRepository</eventIdentifierType>
			<eventIdentifierValue>validacion</eventIdentifierValue>
		</eventIdentifier>
		<eventType>validation</eventType>
		<eventDateTime>2013-03-10T12:00:00</eventDateTime>
		<eventDetail>jhove1_9a1</eventDetail>
		<eventOutcomeInformation>
			<eventOutcome>successful</eventOutcome>
			<eventOutcomeDetail>
				<eventOutcomeDetailNote>Well-formed and valid</eventOutcomeDe-
				tailNote>
			</eventOutcomeDetail>
		</eventOutcomeInformation>
	</event>
</premis>
```

SOLUCIONES APLICADAS

Problemática por sectores

Los distintos tipos de objetos digitales y los sectores de actividad en los que son usados determinan que nos encontremos con distintos tipos de problemáticas de preservación digital. Aunque todos los objetos digitales están sometidos a similares problemas de obsolescencia tecnológica, la importancia que se da a sus propiedades no es la misma, como tampoco es la misma la incidencia que pueden tener los distintos aspectos organizativos, económicos, legales y tecnológicos. Unos pocos ejemplos nos lo pondrán claro: mientras que en la preservación de las películas cinematográficas es clave la salvaguarda de los derechos de propiedad intelectual, en la preservación de datos científicos es suficiente con la conservación de su autoría, dándose, por el contrario, más importancia a facilitar su difusión. De la misma manera la gestión de los datos médicos está totalmente regulada (cuáles son, quién tiene acceso a los mismos, su periodo de retención...), cosa que no ocurre en absoluto con los comentarios entrados en las redes sociales.

Vamos a repasar a continuación las principales características que presenta la preservación en determinados sectores, sin intención de ser exhaustivos.

Sector editorial

La industria editorial ha sido una de las primeras en abordar y en algunos casos solucionar el problema de la preservación de su producción; ello se ha producido por varias razones:

- Los editores han sido conscientes de que su futuro depende de la producción digital, de forma que una posible pérdida de datos podría suponer la caída de un sello editorial y su desaparición como empresa.
- En aquellos productos en los que la contratación se produce bajo la forma de licencias de uso anuales (llámense subscripciones o con otro nombre),

los clientes han pedido garantías de que la conservación y por tanto la disponibilidad para su consulta están garantizadas a largo plazo.
- El alto nivel de estandarización de los productos (concepto de libro, de revista, de artículo, de base de datos) ha permitido usar un bajo número de formatos técnicos y de metadatos, fácilmente automatizables.

La innovación en preservación se inició con la producción de revistas científicas electrónicas en la segunda mitad de la década de 1990, confluyendo en ella los intereses complementarios de los editores y los clientes, especialmente las bibliotecas universitarias. Como resultado de la búsqueda de un equilibrio entre la protección de los derechos de explotación retenidos por los editores, la disponibilidad de la consulta por parte de los clientes y minimizar los elevados costes de mantener la infraestructura técnica para preservar millones de artículos de revista, se ha llegado a un punto en el que la preservación digital de las revistas electrónicas ha quedado delegada en manos de unos pocos grandes depósitos de conservación:

- *Depósitos a cargo de grandes instituciones bibliotecarias*, como Koninklijke Bibliotheek e-Depot, la biblioteca nacional de los Países Bajos, la British Library, el Electronic Journal Center del consorcio OhioLINK, Kopal / Die Deutsche Bibliothek, Los Alamos National Laboratory Research Library (LANL) y el depósito PANDORA de la National Library of Australia. En todos los casos se trata de depósitos centralizados, altamente tecnificados y que se alimentan gracias a los acuerdos que en cada caso se han establecido con los editores.
- *Depósitos privados*, como Portico (http://www.portico.org/), en Estados Unidos, que a cambio de un pago anual tanto de editores como de bibliotecas clientes, se hace cargo de la custodia de los ficheros proporcionados por las editoriales y de facilitar su consulta en caso de un fallo en los sistemas informáticos de estas o de su desaparición como empresa.
- *Depósitos cooperativos*, como LOCKSS (http://www.lockss.org/). LOCKSS es un software desarrollado por la Universidad de Stanford a partir del año 2001. Técnicamente se basa en los sistemas de comunicaciones *peer to peer* y su objetivo es crear redes de preservación distribuida entre instituciones. Cada institución adherida a una red LOCKSS instala el software en su servidor de preservación. Este monitoriza los ficheros almacenados y comprueba su integridad comparándolos con objetos

idénticos almacenados en las otras instituciones. En caso de sufrir alguna pérdida o deterioro de los ficheros, el sistema automáticamente reconstruye la colección a partir de duplicados que recibe del resto de la red. Se trata, por tanto, de un sistema autónomo de ayuda entre depósitos replicados total o parcialmente, sin que sea necesaria la existencia de un nodo central. Este tipo de arquitectura es más barata que los grandes sistemas centralizados y parece más robusta ante incidencias locales. Por contra, requiere que todos los participantes en la red sean escrupulosos con sus obligaciones técnicas y cumplan con las estipulaciones legales.

- *Soluciones propias de cada empresa editorial.* Los grandes grupos editoriales se han dotados de complejos sistemas informáticos que aseguren la permanencia de su producción digital, que es la garantía de su futuro como empresas. En muchos casos han decidido reducir aún más los riesgos apostando por adherirse de forma adicional a alguno de los sistemas anteriormente presentados, tanto por convicción técnica como para demostrar ante sus clientes su compromiso con la preservación de los datos. Es un buen ejemplo de ello la política de Elsevier, que además de conservar sus ficheros en su propio sistema de preservación alojado en un antiguo búnker antiatómico del ejército norteamericano, en Ohio (Estados Unidos), también los deposita en el e-Depot de los Países Bajos, en la red CLOCKSS (un subconjunto más cerrado de LOCKSS) y en el depósito privado Portico.

Con la aparición de los libros electrónicos, los editores están siguiendo las medidas ya ensayadas con las revistas, de forma que observamos que la mayoría de sistemas mencionados previamente para ellas se están ampliando para albergar también los ficheros de los libros electrónicos.

Estas soluciones están constituidas por centros de datos seguros con sistemas automatizados de ingestión. Los datos son entregados por los editores en formatos establecidos, básicamente estructuras XML y ficheros PDF, y usando esquemas de ingestión como NISO Journal Article Tag Suite. Después de la validación de los datos, se generan metadatos administrativos PREMIS y los ficheros se almacenan con control de integridad a nivel de bit.

Cine, vídeo y televisión

Asegurar la permanencia de las grabaciones cinematográficas digitales es un difícil reto provocado sobre todo por el volumen de los datos a preservar y los costes que ello supone. Desde el año 2007 la Academy of Motion Picture Arts and Sciences (la Academia de los premios Oscar) ha publicado diversos estudios en los que se cuantifica el cambio que supone el paso del cine analógico al cine digital. Una película digital de 2 horas de duración ocupa unos 50 TB en alta definición, que se han de multiplicar por las distintas versiones que se generan (por idioma, por país) más el almacenamiento de todas las grabaciones no editadas (los llamados «crudos» de edición). El coste anual de su preservación se multiplica por 12 en el más favorable de los casos respecto a la antigua preservación de las bobinas de celuloide; eso sin contar que se trata de una operación mucho más complicada a nivel técnico debido a la falta de unos estándares claros y a la necesidad de migraciones periódicas. Aunque es cierto que la tecnología digital favorece enormemente la reutilización de las grabaciones, los costes actuales de su almacenamiento parecen inasumibles incluso para las grandes productoras de Hollywood.

La situación en las grabaciones televisivas es algo distinta. Aunque aquí el volumen de horas producidas es aún mucho mayor, se ve compensado por unos requerimientos menores de calidad de los formatos, lo cual favorece la utilización de sistemas de compresión de datos que disminuyen el volumen de datos a almacenar. Por otro lado, la producción está más fragmentada entre cadenas públicas y privadas y productoras independientes, cada una con sus propios derechos de explotación sobre las obras, lo cual complica tomar soluciones homogéneas sobre los fondos.

Mención aparte merece la preservación de los archivos de las televisiones locales, que en muchos casos han cerrado por falta de viabilidad económica o de las frecuencias legales para emitir. Algunos de estos fondos televisivos están llegando a los archivos históricos públicos mientras que otros aún permanecen en manos de las productoras. En todos los casos se trata de volúmenes de miles de horas de producción propia que ocupan decenas de TB. Es esencial realizar una buena elección del formato de almacenamiento según sea la previsión de su explotación futura (dar prioridad a la reedición o, por el

contrario, a la difusión), mantener más de una copia y estar atentos a las futuras necesidades de migración de formatos.

Aún es más difícil la situación de grabaciones distintas a las del cine y la televisión, como pueden ser las de cine *amateur*, las grabaciones publicitarias y las de tipo doméstico. Se trata de fondos técnicamente muy heterogéneos pero que presentan un gran interés documental a largo plazo, por lo que su preservación ha de merecer la atención de las instituciones públicas. Cuando así se produce, pueden ser importantes los problemas legales para disponer de plenos derechos para su transformación digital y posterior uso.

Redes sociales

En las redes sociales como Facebook y Twitter se desarrolla una parte importante de la interacción entre las personas en la actualidad; por tanto, su conservación y consulta puede ser imprescindible en el futuro para entender las sociedades presentes. Pero esta preservación no es fácil. Por una parte, estas redes sociales son propiedad de empresas privadas, la mayoría de ellas de carácter multinacional y con su sede corporativa en el extranjero. Asimismo la preservación a largo plazo no está entre los objetivos de estas empresas, pues no la ven como una fuente de ingresos y sí de gastos (una cosa distinta es la explotación a corto y medio plazo de los registros de hábitos y usos de sus usuarios con fines publicitarios). Otra barrera puede ser la propiedad de estos contenidos: según los acuerdos de adhesión que los usuarios han de aceptar obligatoriamente al acceder a una red social, normalmente todos sus datos (incluyendo las fotos que han subido y los textos que han escrito) pasan a ser propiedad de la empresa que sostiene la red social; ello no impide que los usuarios, que son los autores de estos contenidos, no puedan reclamar algún tipo de derecho sobre los mismos, como los de autoría o de protección de la imagen y la intimidad.

Todo ello presenta un panorama en el que ni las empresas propietarias ni en muchos casos los usuarios son agentes activos en la salvaguarda de los contenidos de estas redes. Una actitud distinta es la de empresas y organizaciones con presencia en la red o bien instituciones que velan por la memoria a largo plazo. Algunas de las primeras desean conservar la interacción que han generado sus perfiles, como es el caso de campañas de *marketing* en Facebook y, por razones distintas, de cuentas gubernamentales en Twitter. En el segundo

grupo, con una visión de recopilación de la memoria histórica, se enmarca el acuerdo de la Library of Congress con Twitter, firmado en el año 2010, para preservar una copia de todos los mensajes que se van cruzando en esta red social. Hay que destacar que Twitter es, con diferencia, la empresa que más facilidades ofrece para la captura y preservación de sus contenidos, incluyendo la disponibilidad de una API que permite su programación.

En paralelo, cada vez aparecen más alternativas de software para realizar la captura de las transacciones que se producen en las redes sociales, normalmente la captura de los movimientos en una cuenta o conjunto de cuentas. Ello no nos ha de llevar a olvidar los problemas jurídicos que permanecen en la acción de captura y, aún más, en su posible consulta posterior, pues al capturar una cuenta propia también estamos capturando sin autorización los contenidos que otras personas han escrito o enlazado.

Datos científicos, *data management plans* y *big data*

Un problema nuevo, que justamente está apareciendo ahora, es el del almacenamiento y la preservación a largo plazo de grandes volúmenes de datos científicos. Cada vez disponemos de más aparatos de captación automática de datos (radares, radiotelescopios, boyas, sondas, videocámaras, sensores, estaciones meteorológicas, medidores de tránsito) que pueden llegar a recoger grandes volúmenes de información en bruto, que en muchos casos se han de procesar casi en tiempo real justamente para sacar provecho de los mismos, y que a veces también se han de poner en relación con los datos que ya se poseían con anterioridad. Además se ha de permitir que investigadores (normalmente desde un acceso remoto) hagan minería de datos con los mismos, todo ello asegurando su preservación a largo plazo. A menudo se trata de proyectos multinacionales o gestionados por grandes centros de investigación en los que se rompen los compartimentos tradicionales entre datos de gestión, datos archivados o semiactivos, copias de seguridad y datos históricos, categorías que, por ejemplo, aún están bien delimitadas cuando nos referimos a datos contables, por ejemplo el presupuesto anual de una administración pública. Algunos ejemplos son los datos científicos recogidos por: CERN, Max Planck Institutes, el radiotelescopio ALMA, WorldWide Protein Data Bank, UK Solar System Data Centre, US Geological Survey e ICSU World Data Center Climate.

Estamos ante bases de datos que se miden por decenas o cientos de TB de las que ya no es posible realizar copias de seguridad de la forma tradicional, pues no existen sistemas con esta capacidad. La preservación se basa en mantener sistemas informáticos replicados, con alta capacidad de crecimiento y con un estudiado diseño que permita la rápida recuperación de los datos necesarios. En estos sistemas los costes de ingestión son reducidos, pues llegan de forma automatizada, pero se disparan los de almacenamiento, al depender de hardware de alto rendimiento.

Una situación distinta es la que presenta la preservación de los datos recopilados en el transcurso de proyectos de investigación de menor entidad. En Estados Unidos y en Reino Unido, encabezando un movimiento internacional que no ha hecho más que empezar, los organismos públicos de financiación de la investigación (la National Science Foundation y los distintos *council*, respectivamente) ya están obligando a los grupos receptores de ayudas a presentar y cumplir con un plan de gestión de datos (*data management plan*) que asegure su correcta gestión, preservación y posible reutilización a largo plazo. Las universidades y los centros de investigación están creando servicios de apoyo a esta ingente tarea, a menudo a través de la biblioteca (se considera que se trata de una extensión de los servicios ya existentes de repositorios institucionales de producción científica) o mediante centros especializados de depósito, como el ICPSR (Inter-University Consortium for Political and Social Research) de la Universidad de Michigan, o el AHDS (Arts and Humanities Data Service) del JISC, en Reino Unido.

La preservación de estos datos de investigación puede ser compleja por distintas razones: muchas veces se tratará de datos mal documentados, formados por un gran volumen de ficheros que además pueden estar en formatos muy especializados, y pueden contener datos sensibles (pensemos en las encuestas y en las pruebas clínicas). Por ello se cree que a menudo no será posible realizar un tratamiento pormenorizado de los ficheros, incluyendo con ello su descripción, y se habrá de optar por soluciones pragmáticas como su almacenamiento en forma de imagen de disco al que se habrá de acceder bajo un entorno virtual.

Una problemática parecida a la de los grandes volúmenes de datos científicos es la del *big data*. Algunas instituciones, incluyendo archivos y bibliotecas, ya están almacenando cientos de TB y millones de ficheros; además

de la problemática y de las soluciones propias que demanden estos tipos de datos, su gran volumen ya es un problema en sí mismo. Un ejemplo lo tenemos en NARA, los archivos nacionales de Estados Unidos: su volumen y, lo que es más temible, su proyección de crecimiento futuro, les están obligando a ensayar alternativas de gestión y de almacenamiento no solo novedosas sino en algunos casos aparentemente heterodoxas. Su sistema de preservación digital fue creado bajo una filosofía distribuida: los documentos se pueden ingresar y consultar desde cualquier archivo de la red federal y los ficheros correspondientes se pueden encontrar almacenados en cualquier ubicación. No solamente no existe un almacenamiento centralizado sino que se está probando su gestión en la nube, de forma que actualmente una parte o una copia de los ficheros se encuentra en centros de datos propios, otra alojada en centros de datos de universidades o de investigación (como el San Diego Supercomputing Center) y otra subcontratada en los almacenamientos privados de Amazon.

Correo electrónico

Los mensajes de correo electrónico son uno de los tipos de documentos digitales en los que existe más interés para que sean conservados a largo plazo, tanto por parte de empresas privadas (pueden ser la constatación de una operación comercial o de la toma de decisiones corporativas), como de administraciones públicas (el correo electrónico ya ha sido equiparado a registros públicos en muchos países), como de particulares (son la base de los nuevos archivos personales).

La problemática se centra en la selección: ¿qué mensajes tienen interés para ser conservados? ¿Es viable acometer la selección de los mensajes dado su gran volumen? ¿Qué es más barato: no seleccionar aunque ello comporte conservar mensajes inútiles, o bien invertir en operaciones de selección para luego ahorrar en almacenamiento? La respuesta de momento no está clara o no es la misma en todas las situaciones. A nivel tecnológico existen diversas soluciones para gestionar el correo personal o el corporativo (incluyendo el filtrado y la consolidación de ficheros adjuntos duplicados). En archivos históricos una tendencia es la fijación de los mensajes en ficheros PDF, con lo que se evita su manipulación y se refuerza su consideración de documento oficial; a la vez, los valores de cabecera (origen, destinatario, tema, fecha...) se

convierten en metadatos que, una vez ingresados en una base de datos, van a permitir una fácil búsqueda y recuperación de los mensajes.

Web

La web y, más específicamente, los sitios web, fueron uno de los primeros tipos de objetos digitales a los que se prestó una gran atención para encontrar soluciones a su preservación. No olvidemos que la web ejemplifica desde su aparición a principios de la década de 1990 las transformaciones que están sufriendo los medios de comunicación y, a pesar de ello, se trata de un sistema altamente efímero, en el que cada nueva aportación a menudo implica la desaparición de la anterior, como es el caso de las actualizaciones de las páginas web.

Muchas legislaciones nacionales ya consideran los sitios web sujetos al depósito legal, lo que se traduce en la recolección de la producción web nacional por parte de las bibliotecas nacionales. El desarrollo informático para hacer que esto sea posible fue desarrollado a mediados de la década de 1990 por Internet Archive (http://archive.org), una fundación norteamericana sin ánimo de lucro que desde el año 1996 está recolectando, preservando y dando acceso a buena parte de los sitios web públicos de todo el mundo. Para llevar a cabo este objetivo desarrolló un nuevo software de búsqueda (Heritrix) y de organización y consulta de los contenidos (NutchWAX y WERA), así como un formato de ficheros para archivo de páginas web (ARC). Todas estas herramientas son de distribución gratuita a través de International Internet Preservation Consortium (IIPC; http://netpreserve.org/) y son las que actualmente están utilizando la mayoría de bibliotecas nacionales y otro tipo de instituciones que preservan la web de un territorio o de otro ámbito específico. Otras instituciones han preferido no crear y mantener una infraestructura propia y, en cambio, subcontratar estas tareas a la fundación Internet Archive, mediante su servicio comercial Archive-It (http://www.archive-it.org/).

Si se trata de capturar y conservar una web corporativa, normalmente por parte de la propia organización para mantener un registro de su presencia externa, existen distintos programas, como HTTrack (http://www.httrack.com/). La operación normalmente consiste en la toma de instantáneas de la web en momentos determinados o con cierta periodicidad; los ficheros capturados (con el nivel de profundidad de los enlaces que se haya

configurado) son guardados en un sistema de almacenamiento (como un disco duro o un DVD) que permite la consulta fuera de línea, aun cuando los ficheros y los enlaces originales ya hayan desaparecido.

Arte

Una parte del arte contemporáneo se ejecuta y se muestra por medios electrónicos y en casos extremos su razón artística es justamente la explotación de las características de los dispositivos digitales, como ocurre en el llamado *media art* y, aún más, en el *net.art*, con acciones basadas en la interacción con ordenadores o en su visionado por web.

Estas manifestaciones artísticas presentan una doble problemática desde el punto de vista de su preservación a largo plazo: por un lado, se basan y dependen de tecnologías y dispositivos efímeros en su propia naturaleza técnica, que acentúan la problemática de la conservación inherente a todas las obras artísticas; por otro lado, se presenta el problema de su exhibición futura, cuando el entorno tecnológico será otro al imaginado por el artista. Ante esta última situación algunos artistas reaccionan exigiendo la conservación estricta del entorno original (lo contrario lo consideran como una manipulación de su obra), mientras que para otros artistas la evolución tecnológica (incluyendo los cambios que conlleva en su obra) forma parte de su interacción con el público y con el medio.

Los museos de arte contemporáneo y los grandes coleccionistas de obras de arte se están enfrentando a esta problemática aunque no existen recetas genéricas. Parece que la emulación de los entornos informáticos originales puede ser una buena solución en muchos casos, pero siempre está sometida a la visión subjetiva que tengan los propios autores. Grandes museos y organizaciones artísticas internacionales como la Fundación Getty y Rhizome ArtBase están trabajando de forma activa en esta línea, hasta el punto de que algunos ya están ampliando sus fondos con obras procedentes directamente de la producción digital. Es el caso del MOMA que está adquiriendo los derechos sobre algunos de los videojuegos más famosos y los está exhibiendo y preservando debido a sus valores artísticos: Pac-Man (1980), Tetris (1984), Myst (1993), SimCity 2000 (1994), The Sims (2000), EVE Online (2003), etc.

Documentación de las administraciones públicas

Las distintas administraciones públicas están gestionando cada vez más documentos electrónicos, resultado tanto de sus procesos internos de informatización como del creciente número de servicios de administración electrónica que ofrecen. La característica fundamental de estos datos es su consideración de documento oficial, con validez jurídica, del que se ha de salvaguardar la integridad y la autenticidad a lo largo del tiempo, una tarea que no es fácil.

La firma electrónica reconocida permite identificar de forma fehaciente el origen de un documento mediante el certificado digital de su creador; también puede asegurar la fecha en la que se creó (sellado de tiempo o *timestamp*), así como su integridad. La firma electrónica ya está ampliamente utilizada en las transacciones entre particulares y también entre particulares y las administraciones públicas, se trata de un instrumento robusto y con respaldo legal, pero que a medio plazo es inconsistente y plantea problemas. El principal de ellos consiste en que las empresas prestadoras de servicios de certificación dejan de respaldar la validez de los certificados digitales al cabo de cierto periodo de tiempo a partir de su emisión, a menudo unos 5 años, con lo que al intentar validar la autenticidad de documentos anteriores en los que estos certificados fueron usados podemos obtener la respuesta de autoría no verificada. Los técnicos en certificación electrónica propugnan que la solución está en resellar (volver a sellar) los documentos con una nueva firma electrónica de forma periódica. Esta solución es ampliamente rechazada por los archiveros por considerarla no sostenible a largo plazo: es complicada técnicamente, es costosa y es difícil asegurar su correcta aplicación en fondos que van creciendo año a año. Los archiveros propugnan usar otras soluciones más sencillas como el uso de apostillas electrónicas o de certificados externos, pues aducen que la necesidad de probar la autenticidad se presenta en el momento de ingresar y de autentificar los documentos en los archivos, y que más adelante simplemente se necesita asegurar su integridad, o sea, que el documento no haya sido alterado.

Pero crear archivos administrativos seguros y fiables desde el punto de vista legal no es fácil para administraciones de tamaño mediano y pequeño si no cuentan con apoyo externo. En estos casos, una solución es la creación de

centros de preservación documental que den servicio a distintas instituciones. Se trata, en definitiva, de servicios de *outsourcing* documental electrónico equivalentes a los que ya existen para el almacenaje de archivos en papel.

Un ejemplo a nivel internacional es el sistema alemán Archisafe (http://www.archisafe.de/). Prácticamente una copia del mismo es el sistema iArxiu (http://www.aoc.cat/Inici/SERVEIS/Gestio-interna/iArxiu) desarrollado por la Agència Catalana de Certificació (CATCert), un organismo autónomo creado por la Generalitat de Catalunya y los municipios. Se trata de un sistema de preservación integral y centralizado (software + hardware + procedimientos) que a nivel operativo funciona bajo la forma de un servicio de almacenamiento remoto de contenidos al que las administraciones, en especial ayuntamientos con baja capacidad tecnológica, remiten ficheros para su custodia y preservación. Un modelo aún más avanzado, al menos en sus prestaciones externas, es Metaposta (https://www.metaposta.com/), del País Vasco. En Metaposta no solo se gestiona información oficial de administraciones públicas sino también de empresas privadas de servicios de interés público, como bancos y suministradoras de electricidad, de forma que los ciudadanos adheridos al servicio cuentan con un buzón personal en el que se almacenan todas las comunicaciones documentales que les atañen.

El modelo impulsado en otros países es la creación de una red integral de sistemas de preservación administrativa que cubra todos los niveles administrativos. Uno de los casos más interesantes es el de Australia, en el que desde los Archivos Nacionales se da soporte (incluyendo software y procedimientos comunes) a los archivos de los estados federales y desde estos a los de los municipios.

Soluciones integrales para archivos y bibliotecas

Es posible que ante una necesidad de preservación digital una empresa o institución cree su propio sistema a partir de software comercial o libre, adaptándolo a sus necesidades. Así se pueden alcanzar soluciones muy bien adaptadas a los requerimientos pero al precio de invertir una gran cantidad de tiempo en análisis y desarrollo. Ante esta situación, son muchos los que prefieren adoptar una solución tecnológica ya preparada, a menudo de carácter comercial, que les permita resolver rápidamente su reto de preservación.

Las soluciones integrales actualmente disponibles no son equivalentes entre sí, pues cada una va dirigida a un segmento distinto de clientes y enfoca la preservación digital de forma diferente. Nos encontramos claramente ante un mercado aún en fase de formación, tanto por parte de la demanda como de la oferta de productos, y ello hace difícil las comparaciones o encontrar una alternativa clara a un sistema determinado. Si intentamos clarificar la oferta existente, nos encontraremos con productos orientados a bibliotecas (Fedora Commons, Rosetta), a archivos (Archivematica, SDB) o de aplicación general (DIAS, HCP, Libsafe). También los hay basados en la nube (Duracloud, Amazon Glacier) o en depósitos seguros externos o *data vaults* (Iron Mountain, Underground Archives, SIAG). A continuación se describen las características de cada uno de estos sistemas.

SDB, Rosetta y DIAS son tres sistemas integrales preparados para ser instalados en grandes organizaciones. SDB (http://www.digital-preservation.com/) está desarrollado por la empresa anglonorteamericana Tessella a partir de su experiencia en la automatización de los archivos nacionales del Reino Unido y de Estados Unidos; se trata, pues, de una solución orientada a grandes archivos, no solo nacionales. Rosetta (http://www.exlibrisgroup.com/category/RosettaOverview), de la empresa israelí ExLibris, se desarrolló a partir de la experiencia de la Biblioteca Nacional de Nueva Zelanda y su mercado está en las bibliotecas nacionales y las grandes universidades. DIAS a su vez es un producto de IBM, creado a partir de su colaboración con la Biblioteca Nacional de los Países Bajos, que se está aplicando en grandes bibliotecas, por ejemplo en Alemania, y en empresas multinacionales.

Fedora Commons y Archivematica también son sistemas integrales de preservación, pero en este caso gratuitos. Fedora Commons (http://fedora-commons.org/) es un software de gestión de repositorios de objetos digitales de cualquier tipo, aunque mayoritariamente se está aplicando para gestionar repositorios institucionales de universidades y bibliotecas digitales avanzadas. Fedora Commons (no confundir con Fedora, una conocida distribución Linux) es muy potente, cuenta con un estricto seguimiento de los estándares y capacidad de integrar nuevos módulos de ampliación, pero justamente por ello tiene fama de ser un software con unos requisitos elevados de configuración y administración, que no se encuentran al alcance de todas las instituciones. Archivematica (https://www.archivematica.org/) es un software libre para

ser aplicado en archivos históricos o administrativos; aunque de momento aún se encuentra en desarrollo, sus versiones preliminares han tenido un gran reconocimiento entre los archiveros.

La alternativa a usar software comercial o libre de preservación digital es el diseño y la implementación de un sistema propio. Esta estrategia, minoritaria en otros ámbitos informáticos, no lo es en el de la preservación digital, pues la ausencia hasta hace poco de productos convincentes llevaba a buscar soluciones propias a las instituciones más avanzadas, algunas de las cuales, como hemos visto, después se han acabado convirtiendo en productos de uso más general.

Un ejemplo de estos desarrollos internos es el sistema SPAR (http://www.bnf.fr/en/professionals/preservation_spar.html), de la Bibliothèque Nationale de France, operativo desde 2010. Se trata de un sistema OAIS completo desarrollado por contratistas franceses y que cuenta con duplicación de infraestructuras entre las sedes de la Biblioteca de París y de Versalles. Otro ejemplo lo tenemos en COFRE (http://www.recercat.cat/handle/2072/97251), puesto en marcha en 2011 por la Biblioteca de Catalunya, en Barcelona. COFRE es un repositorio para preservar los ficheros máster de sus proyectos de digitalización, desarrollado con software libre y que tiene una arquitectura de archivo oscuro, es decir, no está conectado a otros sistemas informáticos y no es accesible para los usuarios. Desde finales del año 2012 también aloja ficheros máster de la biblioteca digital del Ateneu Barcelonès, lo que supone la primera experiencia de preservación digital cooperativa dentro del ámbito español.

Las organizaciones que no deseen o no necesiten implementar la totalidad de las funcionalidades del sistema OAIS, como hacen todos los productos revisados con anterioridad, disponen de diversas soluciones para crear y gestionar un almacenamiento seguro. Hitachi Content Platform (HCP; http://www.hds.com/products/file-and-content/content-platform/), antes conocida como Hitachi Content Archive Platform (HCAP), consiste en un software servidor capaz de controlar el estado de los contenidos de diversos almacenamientos distribuidos de datos, asegurando su integridad, la coherencia entre copias y la aplicación de estrictas políticas de custodia y acceso. HCP se puede aplicar a cualquier tipo de organización que almacene grandes volúmenes de datos, como pueden ser archivos administrativos y hospitales; un ejemplo de

uso es el Archivo del Ayuntamiento de Barcelona. Libsafe, de la empresa española Libnova (http://www.libnova.es/), es una solución software o software + hardware con una orientación similar a HCP, pero especializada para ser aplicada en bibliotecas y archivos, y está siendo usado en la Biblioteca Nacional de España. Otros fabricantes de informática pueden proporcionar soluciones similares a las anteriores, aunque no siempre con una clara orientación a la preservación; es el caso de IBM Information Archive (http://www-03.ibm.com/systems/storage/disk/archive/) y otros

La alternativa al almacenamiento de copias redundantes gestionadas localmente es su almacenamiento en la nube (*cloud storage*). Aquí la empresa líder es Amazon, que ofrece dos productos distintos: S3 (http://aws.amazon.com/s3/) y Glacier (http://aws.amazon.com/glacier/). Amazon S3 permite organizar el almacenamiento de copias de datos en alguno de los distintos centros de datos de que dispone Amazon a lo largo del mundo; de esta forma el almacenamiento se convierte simplemente en un gasto de servicio (SaaS) a pagar mensualmente, no en una inversión en hardware y software que se ha de amortizar a lo largo de varios años. Amazon Glacier es un servicio de almacenamiento de menos prestaciones respecto a S3, especialmente diseñado para la preservación de datos a largo plazo a un coste de solo un 10 % del de S3. Desgraciadamente, en estos momentos tanto S3 como Glacier exigen que el cliente disponga de herramientas propias para controlar los envíos y las consultas de datos en la nube, lo cual no es fácil.

Para paliar este problema una solución puede ser contratar Duracloud (http://www.duracloud.org/), un servicio gestionado por la organización Duraspace que pretende facilitar el uso del almacenamiento en la nube a organizaciones como archivos y bibliotecas. Duracloud es un entorno de software que controla y gestiona el almacenamiento de datos distribuidos en distintos centros de datos de la nube de los proveedores Amazon y SDSC (http://www.sdsc.edu/).

Quienes necesiten cumplir con las máximas exigencias de seguridad de los datos, pueden recurrir a distintas empresas que almacenan los datos digitales en búnkeres antinucleares y blindados contra cualquier intrusión externa. Estos depósitos seguros externos o *data vaults* son bien conocidos y utilizados desde hace décadas, por ejemplo por la industria cinematográfica, que almacena allí sus originales crudos de filmación. El líder mundial en almacena-

miento seguro es la empresa norteamericana Iron Mountain (http://www.ironmountain.com/), pero también podemos destacar otras empresas como la también norteamericana Underground Archives (http://www.uarchives.com/) o la suiza SIAG (http://www.siag.ch/), también conocida como Swiss Fort Knox.

Ejemplos de buenas prácticas

En el apartado previo se han mostrado soluciones técnicas y organizativas que actualmente se están aplicando en distintas partes del mundo. El catálogo presentado es limitado, pero la selección realizada permite comparar alternativas técnicas, aplicadas a entidades públicas y privadas, y a modelos de negocio (la sostenibilidad financiera) diferentes.

A grandes rasgos se observa que existen sistemas de preservación muy centralizados a nivel técnico (e-Depot, Portico, Internet Archive, iArxiu), situación que ellos mismos defienden como una necesidad ante la complejidad técnica de la preservación y las inversiones necesarias para llevarla a cabo. Otros casos presentan sistemas más descentralizados, sea por razones organizativas (NARA), de seguridad (NARA, Elsevier) o por filosofía (LOCKSS).

Los modelos de negocio son variados: asignación del presupuesto público (NARA, e-Depot, LOCKSS, SPAR, COFRE), financiación por los propios usuarios del servicio (Portico, Archive-it, iArxiu, Metaposta) o mantenimiento gracias a recursos propios (Internet Archive, Elsevier). También observamos diferencias en el origen de los datos: internos (NARA, LOCKSS, Elsevier), de terceros (Portico, Archive-it, iArxiu), mixtos (e-Depot) o recopilados (Internet Archive).

Es importante recordar que algunos de estos sistemas han sido desarrollados por instituciones públicas con la colaboración de empresas privadas de ingeniería del software que después han pasado a comercializarlos de forma privada. Así, el desarrollo de e-Depot ha permitido a IBM crear el producto Digital Information Archiving System (DIAS); Tessella vende el Safety Deposit Box (SDB) creado a partir de su experiencia en NARA y en los archivos nacionales del Reino Unido; y Ex Libris vende el sistema Rossetta, basado en el desarrollo creado para la Biblioteca Nacional de Nueva Zelanda. Tampoco podemos olvidar la existencia de numeroso software libre, entre el que destacan los sistemas Xena y Digital Preservation Recorder (DPR) creados

por los archivos nacionales de Australia; DAITTS, creado por el Florida Center for Library Automation; y el software de los proyectos europeos PLANETS, CASPAR y Keep.

ELABORACIÓN DE UN PLAN DE PRESERVACIÓN

En el capítulo anterior hemos visto un catálogo de soluciones de preservación digital que se encuentran en pleno funcionamiento o en fase avanzada de implantación y con anterioridad se han explicado los principales parámetros técnicos en los que se basa la preservación digital. Ante toda esta avalancha de informaciones, algunas ciertamente muy especializadas, cabe preguntarse qué debe hacer una institución determinada que quiera avanzar en este ámbito. ¿Existen prácticas recomendadas? ¿Existes soluciones aplicables de forma genérica o las que hemos visto solo son aplicables a grandes instituciones, con grandes presupuestos? A continuación intentaremos dar respuesta a estos enigmas.

Para emprender un sistema de preservación digital, previamente tiene que formularse una política al respecto y definir unos objetivos: ¿qué vamos a preservar y que no?, ¿con qué finalidad, para que usuarios?, ¿lo vamos a hacer solos o colaborando con otras instituciones o empresas?, ¿preservar forma parte de nuestras responsabilidades o es una actividad voluntaria?, ¿disponemos del conocimiento técnico para hacerlo?, ¿tenemos asegurados los medios económicos para su sostenibilidad a largo plazo?...

Es difícil que un sistema de preservación sea viable si los procedimientos previos de gestión y explotación de documentos y datos no están bien asegurados. Así, una empresa o administración pública primero deberá disponer de un buen sistema de gestión documental o ERM antes de emprender la preservación a largo plazo. Del mismo modo, una biblioteca universitaria primero debe dotarse de un buen repositorio digital y de todos los procedimientos de gestión anexos.

Existe una tendencia equivocada a centrar las preocupaciones en la preservación a largo plazo (décadas) mientras se descuidan las acciones a corto plazo (quinquenio), mucho más viables y urgentes a la vez. En efecto, lo

primero que debería de asegurarse es la fiabilidad y seguridad de los sistemas de gestión presentes, asegurando su robustez ante:

- El crecimiento de los contenidos.
- El incremento en el uso.
- Las incidencias informáticas.
- Los ataques a su seguridad.

Para conseguirlo es necesario actuar en cuatro niveles sucesivos:

- Disponer de sistemas de copias de seguridad fiables, alguna de ellas en localizaciones externas.
- Establecer procedimientos de garantía de la calidad del servicio.
- Realizar auditorías de seguridad. Una buena pauta puede ser seguir las indicaciones establecidas en el Esquema Nacional de Seguridad (ENS).
- Intentar establecer sistemas informáticos redundantes (duplicados), quizás en colaboración con otras instituciones similares.

En cuanto a los datos, es importante normalizar el formato en que estos se crean y almacenan. El objetivo es la reducción del número total de formatos a gestionar y que los finalmente elegidos puedan contar con mejor soporte técnico. Esta práctica está siendo implementada en todos los archivos nacionales que han empezado a gestionar datos electrónicos.

También es importante hablar con los usuarios potenciales (llamados *comunidad designada* en el argot de preservación digital), para conocer de forma exacta cuáles son sus expectativas, qué propiedades de los documentos o de los datos son las más significativas para ellos y cómo esperan interactuar en el futuro con los contenidos. Este conocimiento ha de servir para configurar el sistema de preservación digital de forma que satisfaga el máximo número de estas expectativas. Fijémonos en que, por ejemplo, para una administración pública lo primordial es asegurar la autenticidad y validez jurídica de los datos, aunque no necesariamente su imagen externa; en cambio para un artista de *media art* la forma puede ser crucial y para un artista de *net.art* la forma ha de estar ligada a la posibilidad de poder interactuar con la misma. Cada comunidad de usuarios tiene sus expectativas, sus formatos comunes de ficheros, sus formatos de metadatos, el nivel de complejidad con el que está dispuesta a trabajar, etc. Pensemos en el ejemplo de los formatos gráficos: un

artista de *net.art* puede trabajar con GIF, una biblioteca digital con JPEG, un archivo histórico con TIFF, un fotógrafo profesional con DNG y un observatorio astronómico con FITS; sería una gran equivocación intentar normalizar estos formatos de entrada y migrarlos a otro distinto, los usuarios del presente y del futuro no lo perdonarían.

Por último, a lo largo del texto ya ha quedado patente que la preservación digital es una actividad compleja a nivel técnico y cara a nivel económico. Estos aspectos se han de tener bien presentes en el momento de asumir responsabilidades de preservación; claramente se ha de impulsar que estas sean debidamente ejercidas por aquellas instituciones que tienen la conservación entre sus funciones. En cualquier caso es obligatorio explorar las posibilidades de emprender acciones de forma coordinada o consorciada. Esta es una tendencia claramente visible en Estados Unidos, en Australia y en distintos países europeos, donde se están uniendo esfuerzos y compartiendo recursos para crear sistemas de preservación comunes o bien, a un nivel más simple, para intercambiar copias de seguridad de los datos.

Un ejemplo máximo de esta tendencia es el consorcio HathiTrust (http://www.hathitrust.org/), creado en 2008, que da seguridad a los contenidos de unas 60 instituciones, entre ellas las principales bibliotecas digitales de Estados Unidos, incluyendo Berkeley, Duke, Emory, Harvard, Johns Hopkins, MIT, Michigan, New York, Princeton, Stanford y Yale. A este consorcio también se ha unido la Biblioteca de la Universidad Complutense de Madrid, lo que la ha convertido en su primer socio no norteamericano. Su modelo de funcionamiento se basa en el mantenimiento mancomunado de un centro de datos distribuido entre la Indiana University y la University of Michigan en el que se alberga un duplicado de todos los libros digitalizados por los distintos socios. HathiTrust también mantiene un catálogo abierto con todos sus fondos y permite su consulta según las condiciones de *copyright* de cada obra y el estatus del usuario.

Un segundo ejemplo relevante es el del consorcio MetaArchive Cooperative (http://www.metaarchive.org/), que empezó en el año 2002 como una agrupación de seis bibliotecas universitarias del sudeste de Estados Unidos. En la actualidad está formado por unas cincuenta instituciones de cuatro países, ente las que se encuentra el Consorci de Biblioteques Universitàries de Cata-

lunya (CBUC), de España. A nivel técnico MetaArchive se basa en el mantenimiento de una red LOCKSS, con la que aseguran los objetos digitales de las distintas bibliotecas participantes.

RECAPITULACIÓN

En preservación digital no podemos dejar que los problemas tecnológicos nos hagan olvidar que también es imprescindible resolver los problemas que sin duda encontraremos a nivel organizativo, económico y legal.

OAIS es el modelo funcional que explica el funcionamiento de los grandes sistemas de preservación digital, pero ello no significa que deba ser implementado de forma completa. De hecho, uno de los puntos clave para el éxito de cualquier proyecto es saber adaptar OAIS a las necesidades estrictas de cada caso y reducir su complejidad a la mínima necesaria.

Aún no podemos afirmar que las actividades de preservación digital estén en un estado de madurez y desgraciadamente son mayoría las empresas y las organizaciones que no se han enfrentado a este reto. Posiblemente debido a ello el mercado de las aplicaciones de preservación es pequeño, lo que provoca que no haya madurado y se haya favorecido la aparición de un número mayor de alternativas, mejores y a menor precio. Aun así, es importante empezar a actuar pero recordando que posiblemente nuestra planificación deberá ser revisada a medio plazo para adaptarla a los nuevos desarrollos que aparecerán y estarán a nuestra disposición. Dentro de estos planes de actuación, la cooperación con otras entidades es imprescindible, pues la preservación es un reto demasiado complejo para ser acometido de forma solitaria.

Las actividades de preservación digital son delicadas y en algunos aspectos aún se encuentran en fase de prueba, pero ello no ha de ser excusa para no entrar en ellas. Recordemos que los contenidos digitales son muy frágiles y sin nuestra activa intervención se van a perder sin remisión. Sería una vergüenza que las generaciones contemporáneas, las que se consideran más avanzadas de la historia de la humanidad, no sean capaces de conservar su producción, como sí lo hicieron generaciones anteriores con otras tecnologías.

BIBLIOGRAFÍA

- *Audit and certification of trustworthy digital repositories. ISO 16363:2012* (2011). Washington: Consultative Committee for Space Data Systems. http://public.ccsds.org/publications/archive/652x0m1.pdf
- **Beagrie, Neil** *et al.* (2008). *Digital preservation policies study. Part 1: Final report*. http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf
- **—; Jones, Maggie** (2001). *Preservation Management of Digital Materials: A Handbook*. Digital Preservation Coalition. http://www.dpconline.org/advice/preservationhandbook
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2009). *Sustainable economics for a digital planet: Ensuring long term access to digital information. Final report*. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- Committee on current records in an electronic environment (2005). *Electronic records: a workbook for archivists*. París: International Council on Archives. http://www.wien2004.ica.org/sites/default/files/Study16ENG_5_2.pdf
- **Deegan, Marilyn; Tanner, Simon** (eds.) (2006). *Digital preservation*. Londres: Facet.
- *Digital Preservation Policies. Guidance for archives* (2011). http://www.nationalarchives.gov.uk/documents/information-management/digital-preservation-policies-guidance-draft-v4.2.pdf
- *Directrices para la preservación del patrimonio digital* (2003). París: UNESCO. http://unesdoc.unesco.org/images/0013/001300/130071s.pdf
- **Gilliland-Swetland, Anne J.** (2000). *Enduring paradigm, new opportunities: the value of the archival perspective in the digital environment*. Washington: Council on Library and Information Resources. http://www.clir.org/pubs/reports/pub89/pub89.pdf
- **Harvey, Ross** (2012). *Preserving digital materials*. Berlín: De Gruyter Saur, 2.ª ed.

- **Hoorens, Stijn** *et al.* (2007). *Addressing the uncertain future of preserving the past; towards a robust strategy for digital archiving and preservation*. Santa Monica, Ca.: Rand Corporation. http://rand.org/pubs/technical_reports/2007/RAND_TR510.pdf
- **Keefer, Alice; Gallart, Núria** (2007). *La preservación de recursos digitales: el reto para las bibliotecas del siglo* XXI. Barcelona: UOC.
- **Kirschenbaum, Matthew G.; Ovenden, Richard; Redwine, Gabriela** (2010). *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington: Council on Library and Information Resources. http://www.clir.org/pubs/reports/pub149/pub149.pdf
- *PoWR: the preservation of web resources handbook* (2008). University of London Computer Centre, UKOLN, JISC. http://www.jisc.ac.uk/media/documents/programmes/preservation/powrhandbookv1.pdf
- *PREMIS data dictionary for preservation metadata. Version 2.1* (2008). http://www.loc.gov/standards/premis/v2/premis-2-1.pdf
- *Reference Model for an Open Archival Information System (OAIS). Recommended Practice. Magenta Book* (2012). Washington: Consultative Committee for Space Data Systems (CCSDS), junio. http://public.ccsds.org/publications/archive/650x0m2.pdf
- **Serra Serra, Jordi** (2008). *Los documentos electrónicos. Qué son y cómo se tratan*. Gijón: Trea.
- *The digital dilemma. Strategic issues in archiving and accessing digital motion picture materials* (2007). Academy of Motion Picture Arts and Sciences. http://www.oscars.org/science-technology/council/projects/digitaldilemma/index.html